A Discussion of the Cybernetics of Learning Behaviour

GORDON PASK

System Research Ltd, Richmond Surrey (Great Britain)

CONTENTS

1. Previous models of learning 178
   1.1. Psychological and physiological models 178
   1.2. The approach of cybernetics 178

2. The conception of learning. 179
   2.1. Adaptive behaviour and learning 179
   2.2. The implications of learning 180
   2.3. The criterion of identification 180
   2.4. Restrictions imposed upon mechanisms and models of learning. 180
   2.5. Finite automata 180
   2.6. Need for many representations 181
   2.7. The field of enquiry. 181

3. Experimental verification 182
   3.1. Difficulty over linguistic ideas 182
   3.2. Environments where learning occurs 182
   3.3. Tests for learning behaviour 185
   3.4. Systems capable of learning 186
   3.5. The mechanism of learning 189

4. Evolving organisms 189
   4.1. Evolutionary systems 189
   4.2. Invariance of an organisation 191
   4.3. The origin of the indeterminacy associated with learning. 191
   4.4. Identification of many evolutionary systems in a brain. 192

5. Physical mechanisms 192
   5.1. Particular model 192
   5.2. Internal environment 193
   5.3. Memory systems, learning mechanisms, and the evolving organisms 193
   5.4. The α system 194
   5.5. The β system. 195
   5.6. Identification between α and β systems 196

6. The evolution of organisms 197
   6.1. Organisms as decision makers 197
   6.2. Interaction with evolving organisms. 198
   6.3. A self organising system. 199
   6.4. The mechanism of evolution 200
   6.5. The necessity for evolution 200
   6.6. Learning related to evolution 200

7. Digression into physical assemblies 201
   7.1. Solving an undecidable problem 201
   7.2. Decision in unstable equilibrium 201

References p. 212-214

8. Different kinds of learning 202
  8.1. Forms of learning mechanism 202
  8.2. Relation between organisations 203
Appendix 204
References 212

1. PREVIOUS MODELS OF LEARNING

1.1. Psychological and physiological models

Psychological models of learning behaviour normally describe what changes will occur as a result of a stipulated experience; they consider the 'before' and 'after' stationary states of the organism, and, on the whole, neglect the non stationary transition. It is true that statistical models (Estes, 1950; Bush and Mosteller, 1955; Siegal, 1959) make a broad comment upon the actual process, but their predictive value is confined to parameters such as the number of trials required to reach a given criterion of performance. Few of the psychological models propose a mechanism for learning. Those that do are either testable but restricted, *e.g.* Brown's (1958) model for short term memory, Broadbent's (1958) filter model or, as in the case of Hull (1952) and Eysenck (1956) there is some doubt about the status of the intervening variables that are introduced. Without exception, the testable models are open to the criticism that they deal with learning in very restricted conditions such as a Skinner box (1938) or a maze, and consequently, although they refer to adaptive behaviour amongst a specified set of alternatives their validity as models of learning depends upon the assumption that this facet of behaviour can be regarded as relevant and the rest of it (the maundering of the rat, the idle thoughts of the man) can be neglected. The evidence seems to contradict this supposition and (whilst it is difficult to offer an alternative) it is fair to say that this form of model refers to reinforced selection and completely omits the issue of attention.

Physiological models (Pavlov, 1927; Sherrington, 1948; Hyden, 1960; Bok, 1961) rightly avoid commitment to learning as such and describe some process within the brain. Thus Pavlov was acutely aware of the difficulties of extending his conditioning models beyond the experimentally verified region. In general, with any physiological model, it is necessary to distinguish possibility from actuality with great care. A conditioning mechanism, for example, undoubtedly could account for adaptive behaviour; but that is different from saying it does account for this behaviour in every organism or upon every occasion.

1.2. The approach of cybernetics

Some cybernetic models are derived from a psychological root, for example, Rosenblatt's (1961) perceptron and George's (1961) automata stem largely from Hebb's (1949) theory. Others, such as Grey Walter's (1953) and Angyan's (1958) respective tortoises, have a broader behavioural antecedent.

On the other hand, neurone models, like Harmon's (1961) and Lettvin's (1959), are based upon facts of microscopic physiology and have the same predictive power linked to the same restrictions as an overtly physiological construction.

Next, there are models which start from a few physiological facts such as known characteristics or connectivities of neurones and add to these certain cybernetically plausible assumptions. At a microscopic level, McCulloch's (1960) work is the most explicit case of this technique (though it does not, in fact, refer to adaptation so much as to perception) for its assumptions stem from Boolean Logic (Rashevsky, 1960, describes a number of networks that *are* adaptive). Uttley (1956), using a different set of assumptions, considered the hypothesis that conditional probability computation occurs extensively in the nervous system. At a macroscopic level, Beurle (1954) has constructed a statistical mechanical model involving a population of artificial neurones which has been successfully simulated, whilst Napalkov's (1961) proposals lie between the microscopic and macroscopic extremes.

Cyberneticians are naturally concerned with the logic of large systems and the logical calibre of the learning process. Thus Willis (1959) and Cameron (1960) point out the advantages and limitations of threshold logic. Papert (1960) considers the constraints imposed upon the adaptive process in a wholly arbitrary network, and Ivahnenko (1962) recently published a series of papers reconciling the presently opposed idea of the brain as an undifferentiated fully malleable system and as a well structured device that has a few adaptive parameters. MacKay (1951) has discussed the philosophy of learning as such, the implications of the word and the extent to which learning behaviour *can* be simulated; in addition to which he has proposed a number of brain-like automata. But it is Ashby (1956) who takes the purely cybernetic approach to learning. Physiological mechanisms are shown to be special cases of completely general systems exhibiting principles such as homeostasis and dynamic stability. He considers the behaviour of these systems in different experimental conditions and displays such statements as 'the system learns' or 'the system has a memory' in their true colour as assertions that are made relative to a particular observer.

2. THE CONCEPTION OF LEARNING

2.1. Adaptive behaviour and learning

The adaptive behaviours called 'learning' are inadequately demarcated. As a matter of logic, 'learning' must be held apart from adaptation alone (although adaptation always accompanies 'learning'), for without this distinction, most mechanistic comments on the subject become vacuous. Ashby has shown that any large dynamic assembly must adapt and retain some mark of its previous configuration. Hence, to say that a brain or a large artifact *does* exhibit this behaviour, is an undiscriminating and almost tautologous remark. By the same token, 'memory' (inferred from an observed learning behaviour) must be distinguished from 'retention'.

Another of Ashby's results (1958) has demolished the once popular belief that 'learning' is simply an adaptation *directed* to some objective, like sustenance of the learning organism. This view is far too naive; all stable systems are homeostatic in a tenable environment and all homeostasis can be credited with an objective. If the environment is perturbed in a particularly regular fashion the adaptive changes that maintain the homeostasis are called *habituation* which is a predictable consequence of elaborate structure.

2.2. *The implications of learning*

It is generally agreed that over and above directed adaptation learning entails the construction and appreciation of order amongst objects and their images. An organism that learns becomes able to deal coherently with its environment, or the flux of its autonomous activity, by imposing a plan upon otherwise haphazardly arranged constituents. Minsky and Selfridge (1960), in the idiom of 'artificial intelligence' require generalisation and the development of 'a heuristic' as indices of learning, whilst MacKay (1956) has maintained the very similar view, that learning of any kind entails building an hierarchy of *languages* or systems of signs and relations. The hierarchy is realistic insofar as some translation can take place between these different levels of discourse.

2.3. *The criterion of identification*

For the present purpose we shall concentrate upon the act of identification, or in Carnap's (1947) sense, an interpretation of some nonlogical sign in a calculus in terms of a designatum, as the crucial feature of learning ('translation of a message between languages' and 'identification of a pair of otherwise incomparable sign sequences' are operationally equivalent). Consequently an organism that *learns* must also be able to identify relevant stimuli (or values of attributes in its environment), to bring its attention to bear upon a limited universe of discourse which includes these stimuli and the response alternatives amongst which adaptation occurs. Similarly, it must form and use named 'concepts' or generalisations, the crux of this process being identification between distinct concepts (called 'insight') rather than the aggregation of common features which is a relatively trivial result of constrained adaptation (notice that this kind of 'identification' is a 'translation').

2.4. *Restrictions imposed upon mechanisms and models of learning*

Once a more sophisticated view of learning has been accepted, we are forced to discard most of the neat and familiar representations of mental or neural activity. In particular, the restricted models of metricised information theory are inadequate vehicles for a description of any *mechanism* supposed to account for learning behaviour (Pask, 1958a,b). Of course, there is a sense in which information is communicated and transformed when an organism is learning but as Greene (1961) points out (when he remarks that percepts carry with them overtones of other messages) this is not the sense of any metricised information theory. McCulloch and MacKay (1952) said much the same when they discussed the status of 'signals' in a brain. Information is represented in various ways, for example, in terms of impulse rate and impulse phase relation and neither an impulse nor any other event is unambiguously a signal.

2.5. *Finite automata*

Now information theory is applicable within the framework of any one representation. This representation defines a set of permissible outcomes and messages and codes. Similarly, any one representation defines a set of abstract models which can be identified with a physical brain, by an observer who specifies his state description and forms a *system* in the sense of Ashby. Certain of these systems are *isomorphic* or *homomorphic* images of one another, and *within* such a well related *set* of systems there is a calculus equivalent to the calculus of information theory. A typical abstract model is a McCulloch and Pitts (1947) network of 'artificial neurones', a typical system, its identification with attributes of some part of a brain, and in this case the transformable set of systems is the set of (possibly restricted) finite automata.

But there is no guarantee that information theory or any other calculus is applicable *between* representations (indeed we must assume the contrary).

2.6. Need for many representations

Now we have argued that learning is crucially dependent upon an act of identification, in other words, upon a relation *between* different representations of data. But in this case, the process of learning cannot be described, or at best, can rarely be described, in terms of a single information system. Normally it will be necessary for an observer who investigates learning to reconstruct *his* identification and remake *his* state description whenever the observed organism 'changes its attention', or exhibits 'insightful behaviour' (that is, whenever *it* changes *its* identification).

No *single* system can represent the mechanism that underlies learning. No finite automaton can be held to learn. A search for *the* mechanism of learning is either unduly parochial or misconceived for it is the activity of many systems, not one, that is correlated with learning behaviour (or, we can say, it is a single self organising system, as the phrase is used by Von Foerster, 1960, of which the separate systems are components, related by the appearance of learning and possibly in no other way).

2.7. The field of enquiry

It is important to recognize that these comments apply to a field of enquiry, namely 'learning behaviour' and not to all facets of mental or neural activity. Thus Harmon (1959) is on firm enough ground when he says that impulses are the minimal components of messages in *the* language of neurones, simply because Harmon is not directly concerned with learning and is consequently able to consider events in a *single* language system. But since we *are* concerned with learning, the pertinent enquiries refer to many language systems (and their structure is changed by the process being investigated). Thus we cannot afford such elegant and convenient models nor at the present state of knowledge can assertions about learning be reduced, say, by Nagel's (1949) procedure, to expressions in the conceptual framework of neural networks.

On the other hand, it would be just as absurd to suggest that networks and finite automata are valueless in the role of explanatory devices as it would be to suggest that bond models of chemical molecules are valueless explanatory devices in reaction chemistry. For whilst it is legitimate to remark that neurone networks (or bond models) are unable to answer our enquiries concerning the mechanisms of learning (or the dynamics of a reaction) probabilistic combinations of either model have predictive value. Their merit is conceptual simplicity. Their weak point is that given a probabilistic combination which seems to fit the facts, we are left with an indeterminacy* (Pask, 1962; Von Foerster and Pask, 1960/1961), or uncertainty regarding the *form* of organisation ultimately responsible for the observed behaviour.

This indeterminacy comes from speaking in terms of one language about a construct that entails many languages, and it can only be removed by an act of identification, in particular, if the observer translates expressions from the language of network model or bond model descriptions, into the corresponding expressions in a more comprehensive language or *metalanguage*.

3. EXPERIMENTAL VERIFICATION

3.1. Difficulty over linguistic ideas

In order to pass from abstract organisations to physical assemblies, like the brains of men, animals and artifacts, we shall construct several experimental tests for learning. The greatest difficulty at this stage, is confining our discussion to the subject of learning. In an attempt to avoid the whole gamut of linguistic behaviour, I may do violence to many linguistic ideas by inadvertent mistakes or by deliberate but illegitimate specialisation.

3.2. Environments where learning occurs

The primitive constituents of a test situation are an organism (said, after testing, to *learn* or not to learn), its environment of objects and actions and its observer. Depending upon the extent to which he infers similarity with the organism, the observer may produce and respond to the actions and objects that constitute the environment. If the organism is a man, the observer is assumed to infer considerable similarity, if it is an animal or an artifact rather little similarity. But, in any case, the inference is made on grounds that are remote from the experimental evidence.

There is always a sense in which the environment is commonly named by the organism and the observer. (*1*) Given a strong inference of similarity the set of possible objects and actions is taken to be identified in substantially the same way by the observer and the organism, hence to constitute on the one hand, a set of stimuli, or, in the case of verbal discourse of words, on the other hand, a set of nameable response alternatives, which may also be words or sequences of words. (*2*) Lacking a strong inference of similarity the observer must be able, by previous observation or stimulus equivalence investigations, to *describe* the effective stimuli (and place them in correspondence with the objects and actions he perceives), and to describe the response alternatives for the organism. Thus Lettvin *et al.* (1959) described the perceptual attribute space for the frog and described how a frog could respond.

---

* The word 'indeterminacy' is used to distinguish this uncertainty from the more familiar classical uncertainty, given a set of relevant attributes defining a system, regarding the state of the system. Whereas a 'classical' uncertainty is reduced by making more microscopic observation, 'indeterminacy' is not. Although classical uncertainty may often perturb real observations of brains, it has little to do with the immediate discussion.

We thus characterise the test environment by a set X of signs x for objects and actions affecting the organism as its *input*, and a set Y of signs y for responsive actions affecting objects in the environment and which constitutes its *output*. Manifestly, if selections from X and Y are to be signals representing information, neither X nor Y can be unrestricted. Given this, we call $X \cup Y$ the *finite vocabulary* of an *operational language* (that is, a system consisting of the signs in X or Y and relations that appear as invariants of the organism's behaviour), whereby the organism indulges in discourse with its environment. Tentatively, $x \in X$ will be identified with 'subject' words and the output signs $y \in Y$ with 'predicate' words and the organism itself as any operator capable of performing the 'and so' operation of coherent discourse to produce utterances (xy) (in the sense of M. Mastermann, 1958). The least organism of this kind (which features in our formal discussion) must contemplate or take into account at least the possible pairs x, y, of subjects and predicates when selecting some y from Y.

We should emphasize that any construction that involves organisms and operational languages is specified relative to a scientific or descriptive language λ (used and comprehended by a body of observers). There is no guarantee that the structure of the operational languages will be preserved if the descriptive language is changed, say from $\lambda \rightarrow \lambda^*$. It *may* be possible to represent systems of operational languages in a manner that is invariant with transformations $\lambda \rightarrow \lambda^*$. The recent discussions of cybernetic logic by Gunther (1962) indicate that this can be achieved, at any rate in principle. But, whilst admitting that relativism is unsatisfactory, we shall not discuss the issue any further in the present paper.

Now a scientific language is a metalanguage, in terms of which an operational languages is described. We shall concentrate upon those systems constructed in λ, within which it is possible to distinguish more than one operational language.

Let us denote these operational languages α, β, and their finite vocabularies $X^\alpha \cup Y^\alpha, X^\beta \cup Y^\beta$. Let permissible or sensible utterances have the form $U = (x, y,)$ and denote the set of permissible α utterances $U^\alpha \subset X^\alpha \otimes Y^\alpha$ and the set of permissible β utterances $U^\beta \subset X^\beta \otimes Y^\beta$.

It is not too difficult to specify the conditions that must be satisfied if α and β are to exist in our λ description of a system if we conduct the discussion in rather vague terms. Hence, we shall give a broad account and become a little more precise, when we need to be, later (see Appendix).

(1) If α and β are operational languages in λ it must be possible to distinguish (also in the λ description of the system) distinct *organisms* capable of using these languages for communication.

(2) These organisms must belong to at least a couple of different species (say the A species of organisms and the B species of organisms), and communication must *occur* between the members of each species of organisms; being characterised in the case of α by utterances $u \in U^\alpha$ and in the case of β by utterances $u \in U^\beta$.

(3) In order that α and β be distinct, the A system of communication must be distinct from the B system. The distinction could be secured in various ways of which only one will be cited.

(4) (I) There must be at least some signs, used in the α discourse, which do not

appear in the β discourse and *vice versa*, hence $(X^{\alpha} \cup Y^{\alpha}) \cap (X^{\beta} \cup Y^{\beta}) \subset (X^{\alpha} \cup Y^{\alpha}) \cup (X^{\beta} \cup Y^{\beta})$. (II) It will be convenient to assume that U^{α} and U^{β} are disjoint. Thus $[U^{\alpha} \cap U^{\beta}] = 0$ where 0 denotes the empty set. (III) The intention behind (II) is that U^{α} and U^{β} are sets of utterances which enjoy the property, for any $u_0 \in U^{\alpha}$ that 'u_0 means the same thing to all organisms of the A species' and similarly for any $u_1 \in U^{\beta}$ that 'u_1 means the same thing to all organisms of the B species of organism'. (We shall make the relation of 'meaning the same thing' more precise, later. For the moment, assume that a given utterance 'means the same thing' to a pair of organisms if, when received in the same conditions, it induces the same behaviour). (IV) Hence $[U^{\alpha} \cap U^{\beta}] = 0$ implies that there is no u which means the same thing to *all A* species organisms and *all B* species organisms.

(5) This condition (which entails complete lack of relation between α and β with respect to universally interpreted signs) is a shade more restrictive than necessary, although it will be convenient to adopt it. But in any case we should insist that α and β are neither isomorphic nor homomorphic images of one another when the 'and so' precedance of x, y, is interpreted as a *composition.*

(6) However, there must be a possibility of 'translation' between α and β or *vice versa.* We shall interpret a translation as a special transformation that has the index of the particular organism that performs this translation as its parameter. For this purpose, we introduce the concept of a limited area of 'meaning'. An α sign may mean the same thing to one member of the A species of organism as some β sign or some set or some sequence of β signs (but this cannot be a universal implication, true for all members of the A species of organism). Similarly, translation can be employed by B organisms to interpret α signs.

We shall argue that the possibility of 'translation' rests upon a similarity between the A species organisms and the B species organisms which exists even though they use distinct communication systems. (In our discussion we shall be chiefly concerned with similarities of physical structure or similarities in the relation between these organisms and a common environment, but there are many other possibilities. Thus a French and an Italian enjoy similar cultural backgrounds, and their discourse has a similar semantic content.)

(7) The α operational language and the β operational language remain unrelated with respect to the *universally* true assertions of the participating organisms. But they are, of course, perfectly well related from the viewpoint either of the λ observer or a particular organism that *identifies* or *translates* a set of utterances, hence, according to our criterion, an organism that *learns.*

The possible relations α, β and β, α are (1) that α and β characterise communication systems capable of sustaining the same level of discourse or (2) that α is a metalanguage with reference to the operational language β (and is thus capable of describing some aspect of β) or, *vice versa*, that β is a metalanguage with reference to α. If α is a metalanguage relative to β (or *vice versa*) the functions determining α discourse, are of higher *order* and propositions in α are of a higher logical *type* than the corresponding logical functions and propositions in β; the hierarchy of orders or types being determined by elementary propositions in λ. To elaborate the point, α

signs which might, for example, include 'furniture' cannot be derived from β signs, such as 'table' or 'chair'. To say that 'furniture' cannot be derived from expressions in β means (a) that the category of furniture regarded as a *disjunctive* concept in β must be defined ostensively by exemplars. But the choice of these exemplars is unlimited (we can only specify furniture as the particular tables and chairs and other objects pointed out by a β organism). (b) If the category 'furniture' is regarded as a conjunctive concept entailing those attributes common to tables and chairs, the choice of attributes is unlimited (and we can only specify the particular subset selected by this β organism).

(8) This organism specificity suggests that whatever the logical status of α and β (from the λ observer's point of view) any participant B organism, performing a translation from α to β must use β as a metalanguage. For this B organism must construct a description, based upon its experience, of the usage of α terms. Indeed we shall argue that it must construct within itself an image of an A organism and look at the environment in an A-like representation.

In the limiting case, which we shall examine in connection with the *mechanism* of learning, the analogues of A and B organisms have a repertoire of utterances which completely specify their behaviours. These organisms could be manifest in no other way, and in order to exist, they must make precisely these utterances. A complete α to β translation performed by a B organism thus amounts to constructing an informational replica of an A organism in the medium of the B organism concerned. In other words, an $\alpha \rightarrow \beta$ translation amounts to *reproducing* an A organism. We shall return to this point in a moment; its immediate relevance is that it demonstrates an identity between the chosen criterion of learning and ideas which appear in Wiener's (1961) account of the reproductive process, which he is concerned with developing at the moment (Wiener, 1962). Specifically we shall identify learning with a cybernetic property of organisations that reproduce or evolve in media like brains and social aggregates and will argue that at least in certain special cases (Wiener's concept applies to the general case of the 'Black and White Box' reproductive system) the fact of evolution (which entails some reproductive mechanism) will also entail the manifestation of learning behaviour.

3.3. Tests for learning behaviour

Learning occurs insofar as the system said to learn, is able to identify utterances or signs in its operational language with novel *designata*; if these *designata* are utterances in a different operational language, the organism can learn insofar as it can perform a translation.

Tests of this ability may involve either control or communication and each kind of test is fairly commonly used. So far as *control* is concerned, the experimenter specifies and the subject agrees or is known to accept an *objective* which can only be realised either (I) by recognising attributes of events or patterns of distinct stimuli which have not been encountered upon previous occasions (these are commonly regarded as tests of perception or abstraction), or (II) by constructing a novel categorisation of objects in the environment (these are grouping tests). In either case the subject is required to identify signs in his own language with signs in the language of a feedback loop, stabilising the desired objective, which the experimenter has contrived in the environment. Of course, only a particular form of learning is detected, namely a form which entails order relations that the experimenter can appreciate himself.

So far as communication is concerned the test situation is normally an interview (a special case is considered by John Clark). We assume that any pair of men have different operational languages or equivalently different 'ways of thinking' which they prefer to use. (The test is rarely applied to animals because of obvious but not insurmountable difficulties.) In conversation, 'translation' occurs in the sense that the compromise needed to maintain coherent discourse about an unfamiliar topic involves the construction of a language or 'way of thinking' that is commonly intelligible to the participants. Hence the interviewer who is anxious to evaluate learning endeavours to estimate the other participants' ability to communicate in this situation, discounting, so far as possible, the effect of extraneous variables, such as 'diffidence' or emotional bias. The disadvantage of this method is that the observer who conducts the test is, himself, a participant and his evaluation must be translated into terms of λ, and a rather similar criticism applies to the alternative technique of observing a man who is interacting with a group of other men (to assess the extent to which this individual makes himself understood and manages to understand the point of view adopted by the others).

3.4. Systems capable of learning

In order to exhibit 'learning' as a cybernetic property of organisations, it is necessary to represent the systems we consider in our λ description in sufficiently abstract and general terms. For the present purpose the most convenient representation is Rosen's (1958) wherein a system is conceived as a directed graph, such as illustrated in Fig. 1a.

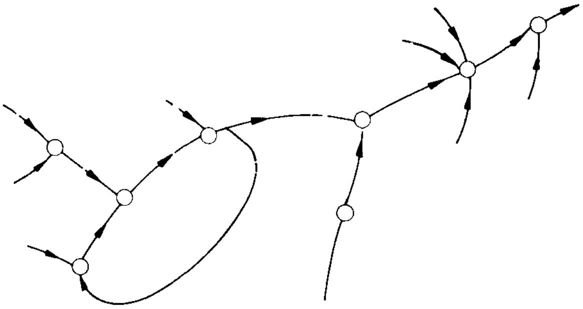


Fig. 1a.

The directed edges of the graph correspond to mappings F, the nodes to the conjunction of the domain of a mapping and the ranges of some other mappings as in the Appendix.

Now this representation can be used to depict either the physical embodiment of a communication system Ξ, in which case the several mappings in the graph represent physically distinct subsystems Ξ_i and coupling functions between these subsystems;

or, alternatively, it may be used to depict the communication system as an information structure Ξ^α or Ξ^β without any particular physical referent; in which case the mappings are transformations of *signs* and the graph determines which sign transforming subsystem Ξ_i^α, Ξ_k^β, can provide an input to another Ξ_i^α, Ξ_l^β. Communication entails the existence of cyclic subgraphs, like *a* and *b* in Fig. 1b.

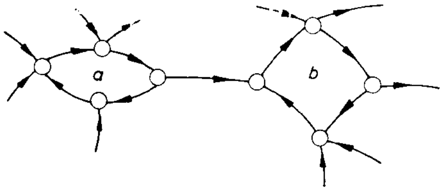


Fig. 1b.

The process of adaptation appears at the level of physical embodiment. It is formalised as change in a parameter φ_i (for subsystem Ξ_i) which selects a particular mapping $f_{\varphi i}$ from a set of possibilities F_i. Since the domain of $f_{\varphi i}$ may be the domain of F_i or a *restriction* of this domain, adaptation not only changes the kind of function that characterises Ξ_i but also the momentary couplings between Ξ_i and the other subsystems (within limits defined by the system graph). All subsystems of interest are adaptive.

To pass from the representation of a physical embodiment of a communication system to the representation of a communication system, it is necessary to abstract special attributes of the physical states which are capable of acting as signs and sequences of which constitute utterances.

We comment that assertions about communication systems and their physical embodiment necessarily are assertions of a different logical type. Now in λ we need to distinguish a pair of communication systems, A and B, mediated by the same physical embodiment; hence we need to consider a pair of abstractions, one of which induces α signs and the other β signs. Within each of the systems A and B communication must occur and the 'signs' must have sensible characteristics, for example, if x is an α sign then it is reasonable to demand that x 'means the same thing' to all of the subsystems A that 'speak' α and which happen to be similarly adapted; but, in order to preserve a distinction between α and β an α sign must not 'mean the same thing' to all B subsystems. Further, from inspection of the directed graph, the calibre of an utterance x, y, is a relation between an individual subsystem and its neighbours (for X_i^α, the set of signs that act as an input to Ξ_i^α is a product set whereas Y_i^α is a factor in a product set) and we insisted, in *3.2.*, that the α and β utterances remain distinct.

Now it appears impossible to offer an acceptably non trivial representation of the phrase 'means the same thing' in terms of the information structures A or B alone. Perhaps the least image which avoids the criticism that we are talking about a calculus rather than a language is 'x means the same thing to Ξ_i and Ξ_j if and only if, when $\varphi_i = \varphi_j$, reception of x by Ξ_i gives rise to the same state of Ξ_i as reception of x by

Ξ_j induces in Ξ_j'. The criterion of similarity is an equivalence relation over the *states* of the physical embodiment. It has the same perhaps arbitrary origin as the partition of states into the states of distinct subsystems and the construction of a directed graph.

It is not too difficult to constrain the system so that $X^\alpha = \cup_i(X_i^\alpha)$, $Y^\alpha = \cup_i(Y_i^\alpha)$ are defined (together with X^β and Y^β), and to ensure that $U^\alpha \cap U^\beta = 0$ *providing*, of course, that the underlying structure of the physical embodiment remains invariant. It is now necessary to enquire what a translation would imply in such a system and how it might be represented.

The particular approach we have chosen is this; within the system as it stands translation cannot occur, but there is no reason why the α and β utterances of coupled subsystems $\Xi_i \cup \Xi_j$ in A, for example, should not 'mean the same thing' as the utterances of coupled subsystems in B, such as $\Xi_k \cup \Xi_l$, for these utterances have the form (u, u), rather than u and if the underlying states (say of Ξ_i) are z in Z_i the states of a coupled subsystem (say of $\Xi_i \cup \Xi_j$), have the form (z, z), in $Z_i \otimes \Xi_j$). Now to introduce a strong coupling, as against the existing concept of communication, is completely arbitrary unless it can be justified apart from our desire for a translation mechanism. We attempt to justify the innovation on the grounds that whenever an act of utterance is *determined* then it does not constitute an act of communication; for no uncertainty regarding the state of the recipient is removed by observing this utterance. Now it is possible to construct a system in which adaptation leads to precisely this deterministic condition so that, in the first place, groups of subsystems are impelled to communicate and then, as a result of communication, become so closely coupled that they can no longer be regarded as separate entities. The construction entails the idea of survival of the fabric or physical embodiment of the Ξ_i which depends, by hypothesis, upon some commodity (represented by an index $\theta(\Xi_i)$). If $\theta(\Xi_i)$ exceeds some arbitrary value θ_0, Ξ_i survives. If not, Ξ_i does not survive and the directed graph Ξ_i^α is not defined; hence any realistic adaptation is such that it maximises $\theta(\Xi_i)$. Let θ be a superadditive measure,

$$\theta(\Xi_i \cup \Xi_j) > \theta(\Xi_i) + \theta(\Xi_j)$$

and cooperation (or communication) is favoured; the rationale of this being that outcomes in $Z_i \otimes Z_j$ which are consistently achieved if and only if the state selections are jointly determined, yield higher values of θ.

The proposed mechanism of translation, and by hypothesis of learning, is of this kind. Given an adaptive system the originally localised subsystems become distributed, *i.e.* Ξ_i and Ξ_j coalesce to form $\Xi_i \cup \Xi_j$. Some α utterances u in use by Ξ_i come to 'mean the same thing' as α utterances (u, u) in use by the distributed system. A similar process occurs in B to produce some β utterances u that are identified with β utterances (u, u) and at this level identification may be (and we assume in a particular system that it is) possible. In this case, because of the original identities between u and (u, u), a translation from $u \in U^\alpha$ into $u \in U^\beta$ is achieved. The underlying physical process which, as it were, 'drives' the system is evolutionary. On consideration it is manifest that some subsystems will fail to survive, hence in reality there must be a replacement or reproductive process. Hence translation involves construction of a metalanguage (since functions defined upon $[\Xi_i \cup \Xi_j] \cup [\Xi_k \cup \Xi_l]$ are functions of the original functions).

We conclude with a few comments:

(I) A subsystem is itself a set of utterances.

(II) To translate these utterances, say from β to α, is a reproductive process.

(III) Rosen (1958) has argued that self reproduction is self contradictory. 'Reproduction' uses a distinguished system such as 'the environment' to act as agent in this process.

(IV) In the present case the 'environment', which is also a prerequisite of efficient reproduction, is represented by the distinct species. Thus Ξ_i and Ξ_j in A jointly act as the environment which is an agent in the reproduction of Ξ_k in B.

3.5. The mechanism of learning

Although we have concentrated upon situations in which learning behaviour is manifest, the argument applies to any organisation whatever. It applies to organisations involving within a man's brain, for example, or in an artifact; and to obtain much the same result as we obtain by experimenting with the exteriorised behaviour.

There are several very elegant models in which organisations evolve. Wiener's mathematical conception has already been cited; Caianiello's (1961) model is ideally adapted to this field of enquiry; in each case stable modes of oscillation are the entities which evolve in the environment of some malleable network, brain or artifact, and which we should identify with evolving *organisms*. Similarly, there are sets of stable modes which correspond to the *species* of evolving organism.

Instead of reviewing these concepts I shall do the less ambitious job of describing a small and in many ways illegitimately simplified model which has been constructed. Tested according to our criteria it does behave as a 'learning mechanism' and could, conceivably, be identified with processes in a real brain. The model arises from the proposition that *learning* is the characteristic behaviour of an evolving species of subsystem in an evolutionary system. A very similar idea was advanced, but differently developed, by Pringle (1951). The particular case we have examined is the evolution of subsystems that are self organising systems: systems with an always positive rate of change of redundancy (Von Foerster, 1960; Beer, 1962). Since this is a *model* rather than an *observed* assembly, it is *constructed* in λ.

4. EVOLVING ORGANISMS

4.1. Evolutionary systems

Evolutionary systems have been considered in detail in other papers (Pask, 1958, 1961a,b). Briefly a system Ξ is evolutionary if:

(*1*) the states of Ξ are partitioned into states of evolving subsystems Ξ_i and of the environment in which they evolve (called the internal environment to avoid confusion);

(*2*) there is a competition between the evolving organisms Ξ_i for some restricted commodity (food, energy, variability, specific metabolite) which is necessary for their survival (we can think of the structural fabric of the organism decaying unless the process is countered by a synthetic activity and suppose that the organisms compete for whatever commodity is needed to maintain this activity);

(*3*) the subsystems Ξ_i of Ξ reproduce their organisation in order to persist. Hence any subsystem or any coupled combination of subsystems is an organisation. We shall not be concerned with specialised reproductive mechanisms like sexual reproduction and it will do no harm to envisage only one reproductive process, namely interaction between a structure which is synthesized from raw material by a controlling process and this controlled activity. Thus activity *induces* the structure, whilst the structure implies the organisation which induces the activity needed in order to build the structure. Hence:

$$\text{STRUCTURE} \leftrightarrows \text{ACTIVITY}$$

(*4*) reproduction entails replication with occasional variation. Commonly there are many structural templets L in a set $\mathscr{L}$ that will induce the *same* activity R or an activity that is *equivalent* in the sense that it belongs to a set $\mathscr{R}$ such that all $R \in \mathscr{R}$ induce some $L \in \mathscr{L}$. Hence the reproductive mechanism is stabilised by a redundancy of structure and function which is represented by a many to many mapping from $\mathscr{L}$ to $\mathscr{R}$ and $\mathscr{R}$ to $\mathscr{L}$. In other words at any moment, t, there will be one particular structure $L_0 \in \mathscr{L}$ which may induce any $R \in \mathscr{R}$ at $t + 1$. Suppose R_0 is induced at $t + 1$, any $L \in \mathscr{L}$ may be induced at $t + 2$.

Thus giving on average:

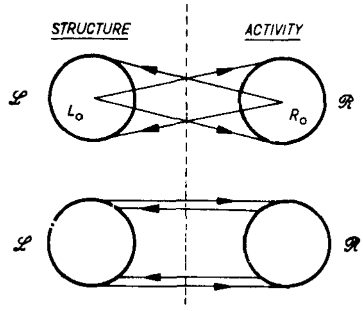


Fig. 2.

The organism that is reproduced is the mapping of $\mathscr{L} \leftrightarrows \mathscr{R}$* and the elements of these sets. It is convenient to denote it G and to write $L(G) \subset \mathscr{L}$ as any *structural* component of G and $R(G) \subset \mathscr{R}$ as any *dynamic* component of G.

However, neither $\mathscr{L}$ nor $\mathscr{R}$ are invariant. To be precise, we should have written $\mathscr{L}(\theta)$ and $\mathscr{R}(\theta)$ where θ is a function of the commodity needed for survival of a structure or the maintenance of activity. Thus for $\theta = 1$ and $\theta = 2$, we might have: Variants arise due to changes in θ (changing $\mathscr{L}(\theta)$ and $\mathscr{R}(\theta)$) or due to aberrant transformation like:

---

* To simulate a viable form of G we choose some characteristic γ, such as a distribution of activity over the entire system, that depends upon the existence of $\mathscr{L}$ and $\mathscr{R}$ and control θ so that survival of an organisation is conditional upon its satisfying γ.

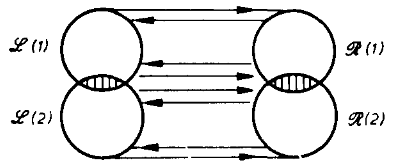


Fig. 3.

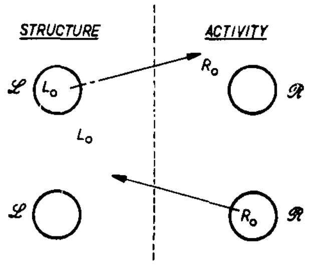


Fig. 4.

or (in a way we shall consider in a moment) due to interaction between a pair of reproductive mechanisms. The important point is that variation is the only *randomness* in the system and it is unrelated to the randomness of dice throwing or haphazard connectivity that appears in a number of adaptive models. The many to many form of $\mathscr{L} \to \mathscr{R}$ and $\mathscr{R} \to \mathscr{L}$ has nothing to do with 'chance' although it may lead an observer to adopt a 'probabilistic' image of the system.

(5) there are constraints in the internal environment that favour the development of specific variants, in this case, of any self organising system;

(6) amongst these are constraints that favour correlated activity amongst a set of organisms realised by a communication between the organisms in this set, essentially, Bonner's (1958) condition for multicellular organisms, and organised populations;

(7) there is a relation of identity between certain organisms defining species of organisms;

(8) the organisms are 'built to survive' that is, they are homeostatic and 'ultrastable' (Ashby, 1956) (they adapt to maintain the possibility of homeostasis in a changing internal environment by maximising θ).

4.2. *Invariance of an organisation*

Identifying a brain with a system such as Ξ, leads to a picture in which the computation performed by the brain is incidental to the basic issue of maintaining the fabric. This is a very reasonable picture for the salient feature of a brain (or any other living assembly) is that unlike a computing machine, it cannot be 'turned off'. If it were 'turned off' by accident, the physical structure would decay and without sufficient metabolic activity it would decay in a less dramatic fashion. The metabolism is simply the energetic side of a reproductive process which preserves organisms within the brain (regarded as the 'internal' environment in which these organisms evolve) or preserves the organisation 'brain' itself.

4.3. *The origin of the indeterminacy associated with learning*

Within this framework the apparent indeterminacy of a learning mechanism has a plausible origin. Assume that the signals that are supposed to represent information in an operational language, say α, are identified in λ with the changes of state constituting the activity component in a reproductive process:

$$\text{STRUCTURE} \leftrightarrows \text{ACTIVITY}$$

The integrity of an informational image depends upon maintaining a distinction between signals and their transformation (or the existence of a transfer function for the entire network or a component, such as a neurone). But the transfer function is determined by the structure. However, the structure is maintained by a reproductive process which entails the flux of activity at least some of which we are anxious to call signals. Hence signals and structure cannot be isolated. Their indices are conjugate variables. Further, the mapping ACTIVITY → STRUCTURE and the mapping STRUCTURE → ACTIVITY are many to many. Thus we could obtain a *partial* but *consistent* form of description of an organism G in λ_u in terms of a probability distribution over the set of structural components of G, but to describe the form of organisation completely it would be necessary to have access to a λ description as in *3.2.* or in the Appendix.

4.4. Identification of many evolutionary systems in a brain

Single evolutionary systems, for which it is not difficult to specify a complete description, can be programmed on a computor (MacKay, 1958; Pask, 1959). But a brain is a physical assembly which is identified with a vast, perhaps an indefinite number of different evolutionary systems. Each one (or, in some cases, each subset of these) can be identified with an operational language, α, β, (in a sense we shall illustrate in a moment). But the evolutionary process in the assembly does not consist of the separate evolution of these systems; on the contrary, interaction occurs between the systems — when the interaction occurs *within* α or within β it has the calibre of information flow, but when it occurs between α and β or β and α, it has the calibre of an *identification*, which, according to the argument of this paper, is the process which distinguishes learning from adaptation.

5. PHYSICAL MECHANISMS

5.1. Particular model

The model we shall discuss has been realised as an artifact (Pask and Lewis, 1962). The logical constituents are a pair of variably coupled evolutionary systems having a common internal environment and identified with operational languages α and β such that α signs are one attribute and β signs another attribute of the same physical event (a unit impulse or an action potential). We assume that an observer is impelled by convention, or sheer conceptual necessity, to identify the internal environment of the system with a rigid framework (which we take to be a network of real or artificial neurones), and is anxious to assign definite locations to the physical events and to regard one or another abstraction of events as a signal. One result of this descriptive method is an indeterminacy, which manifestly has a conceptual rather than a factual origin.

For convenience, I shall tentatively identify parts of the system with parts of a brain, using correspondences that stem from a couple of physiological hypotheses. (In some respects, the working of the system plausibly fits what goes on in a brain.) But the main purpose of identification is to yield a vivid picture. Indeed, (*1*) the system could equally well have been identified at a sociological or a psychological level (*2*). Proof or disproof of the physiological hypotheses would have little effect upon the system, which is a cybernetic model, with a physical cogency that stems from the behaviour of an isomorphic artifact.

5.2. Internal environment

We assume the conditions of *4.2.*, some of them in a specialised form. The internal environment is a network of malleable (or adaptive) neurones denoted a. They may either be real or artificial, but in the latter case their design must embody certain elaborate characteristics of real neurones that will appear in this discussion. Full connectivity would be admissible but impracticable. It is however, important that the neurones be in cyclically fibre connected groups denoted A, wherein neurones compete for a commodity indexed by θ. In a real brain the θ competition may entail a mutual inhibitory mechanism which *does* involve fibres or it may depend upon diffusion of a metabolite, or it may be due to the transfer of metabolites from the glial cells surrounding the cortical neurones.

There are a pair of energetic constraints:

(I) Any region of the network is constrained to manifest a minimum activity, and is physically restricted to a maximum activity dependent upon θ (this condition ensures that neuronal reproductive mechanisms are 'driven' by undifferentiated energy).

(II) Any neurone is required to maintain a minimum metabolic level and is limited to a maximum level dependent upon θ (this condition ensures that intracellular reproductive mechanisms are 'driven' by undifferentiated energy).

Recall that θ is an index of the rate at which structures can be synthesized and maintained. We shall be concerned on the one hand, with relatively macroscopic structures (plastic changes at synaptic connections), and on the other, with microscopic structures (molecular synthesis). Hence the physical restrictions indexed by θ differ greatly from one case to another.

Again we posed in *4.2.* the objective of survival and the possibility of decay. There is no question of *neurones* decaying as such. But there is plenty of evidence that neurones may be more or less reversibly oblivescent, that 'plastic' changes can literally decay, and that a molecular species must be maintained by specific synthesis.

The physical events appearing in the internal environment are impulse patterns of which the spatial attributes are α signs and the temporal attributes, β signs.

5.3. Memory systems, learning mechanisms, and the evolving organisms

First, let us consider the evolutionary system over an interval so brief that it has little chance to evolve. In this *relatively* static picture it will be possible to study the relation between an α system and a β system and the processes of information flow, computation and identification.

Within the static picture evolving organisms which are later to be identified with learning mechanisms appear in a degenerate form as 'memory mechanisms' or systems that reproduce organisations called 'a memory'. Each organism is a subsystem of one or another evolutionary system, the α or the β system (we do not suggest that a real brain has only two systems, probably it has an indefinite number; but two is the least number required to exhibit these principles).

5.4. *The a system*

We assign an operational language a to an evolutionary system which is similar to the resonance structure considered by Pringle (1951). This assignment must, of course, be justified later by examining the relation between this and other systems to ascertain that it does define the relation between a language a and other languages.

The evolving organisms are reproduced by interaction between a spatially distributed pattern of activity in the oscillatory neurone network and a structural templet consisting of plastic modifications of the synapses of some of the neurones a in a group A, for example, the neurones, a^1_1, a^1_2, ... a^1_{n1}, in group A_1. Recalling the driving condition of *5.2.* the activity amongst the neurones of A_1 will depend, in the absence of impulses received from another group, A_2, upon the structural templet of A_1 and any pattern that persists, depends upon a dynamic equilibrium achieved by the reproductive mechanism:

$$\text{Impulse pattern} \leftrightarrows \text{Plastic changes}$$

The organism G, which is being reproduced is a relation between impulse pattern and plastic change, is denoted G^a_i, the superscript for the operational language and the subscript for its identity.

Equivalently, G^a_i, $\equiv \mathscr{L}^a_i \leftrightarrows \mathscr{R}^a_i$. Hence G_i has *dynamic* components R $(G^a_i) \subset \mathscr{R}^a_i$, which are not localised in the network and *structural* components L $(G^a_i) \subset \mathscr{L}^a_i$ which *are* localised within the region A_1 of this network.

Neurones in A_1 are linked by parallel fibres to other regions like the group, $A_2 = a^2_1, a^2_2$, ... a^2_{n2}, and since any dynamic component R(G^a_i) is a pattern of impulses it may be transmitted by this parallel fibre arrangement to A_2.

Now depending upon the state of A_2 (namely which G^a_i are being reproduced at A_2) these impulses may induce some L^a_i and later G^a_i at A_2 because of which we call R(G^a_i) an a sign for G^a_i (it is neither necessarily nor commonly a unique sign for G^a_i). Hence organisms in a may be reproduced at a given location or transmitted through their dynamic component to induce a remote reproductive mechanism, for the same organism. a Signs are the dynamic components of organisms and are the *spatial* pattern *attribute* of certain physical events.

Thus, as a consequence of describing the internal environment in λ as a network and distinguishing an a system:

(I) All a organisms G^a_1 are associated with a reproductive mechanism which, at any instant, has a definite location upon some group, A_1, of neurones.

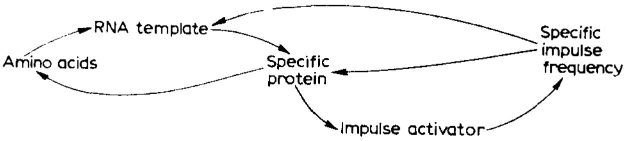
(II) a Signs are dynamic components R(G^a_1).

(III) The set of a signs is determined by the spatial pattern attribute of physical impulse sequences.

(IV) Only a signs have 'meaning' (well defined operational effect in a). They are transmitted by a couplings of appropriate (parallel fibre) capacity.

5.5. The β system

We shall assign an operational language β to an evolutionary system proposed as a memory mechanism by Hyden (1960) as a result of his experiments upon the changes in RNA concentration in stimulated neurones. Similar mechanisms have been suggested by Polonsky (1961) and Barbizet and Albarde (1961). We assume that the frequency (or a comparable attribute) of a temporal sequence of impulses is able to modify some of the RNA molecules produced by a neurone so that a correspondingly modified protein is synthesized. We further suppose that this modified protein (which will only be available in neurones that have previously received a stimulus at this frequency) is specifically sensitive to this stimulation frequency and dissociates, if the frequency occurs upon a subsequent occasion, to yield acetylcholine or some other activator. Hence certain neurones embody 'specific receptors' denoted *b* for specific frequencies of impulse sequences and since the frequency is represented by a modification in RNA synthesis a neurone *a* may include vast numbers of 'specific' receptors, $b(1)$, $b(2)$, ... $b(m)$ for different frequencies. Hence β signs are frequencies or temporal attributes of an impulse sequence, in particular, a β sign is that frequency able to induce a specific receptor *b*. An organism G_i^β in β is maintained by a reproductive mechanism, which recalling the 'driving' condition of *5.2.* is represented by the double cycle:



which might be formulated:

$$G_j^\beta \equiv \mathscr{L}_j^\beta \leftrightarrows \mathscr{L}_j^\beta \leftrightarrows \mathscr{R}_j^\beta$$

which can be reduced to:

$$G_j^\beta \equiv \mathscr{L}_j^\beta \leftrightarrows \mathscr{R}_j^\beta$$

if we keep in mind the fact that the structural components L(G_j^β) are automatically perpetuated (due to the persistence of the RNA modification) and consequently are likely to be distributed (any neurone having a specific receptor will retain it with a minimum of oscillatory impulse reinforcement). Hence the location of the structural component of G_j^β in the network at any instant is likely to be a set of specific receptors,

$$B_j = \left[\overset{1}{b_1}(j), \overset{1}{b_2}(j) \ldots\ldots\ldots\ldots \overset{1}{b_{n1}}(j); \overset{2}{b_1}(j), \overset{2}{b_2}(j) \ldots\ldots \overset{2}{b_{n2}}(j); \ldots \right]$$

which will be very large.

Thus, as a consequence of describing the internal environment in λ as a network and distinguishing a β system:

(I) All β organisms G_i^β are associated with a reproductive mechanism which is distributed (since its structural and its dynamic components are distributed) being located only in a set *B* with specific receptors *b* in most neurones.

(II) β Signs are dynamic components R(G_i^β).

(III) The set of β signs is determined by the frequency attribute of all physical impulse sequences.

(IV) Only β signs have 'meaning' (well defined operational effect) in β. They are transmitted along any β coupling which is any fibre of appropriate (frequency limiting) capacity.

5.6. Identification between α and β systems

The physical events, impulse sequences distinct in λ, have an α and a β aspect. Further, although an α sign has no 'meaning' in β any sequence of α signs implies a frequency of impulses at some point in the network which may be (and by a suitable choice of parameters, almost certainly is) a β sign. Thus, the flux of α signs generates β signs but the β sign (or frequency) generated by a given α activity will normally differ at different points in the network. Conversely, many different modes of α activity will induce the same β sign at a given point.

The signs have lost the property of being signals and bear a closer relation to Greene's *percepts*.

We have argued that a peculiar, location dependent, many to many mapping exists from α to β. Conversely, since any set of frequencies induces a spatial distribution, there is a comparable relation between β and α.

The immediately important point is that our assignment of operational languages, α, β, to a pair of evolutionary systems appears to be sufficiently justified because the relation between α and β is exactly the many to many, λ indeterminate, and organism dependent relationship we discussed in *3.3*. The doubtful part of the argument is the property of organism dependence; to exhibit it notice that the frequency induced (as a β sign) by a sequence of α signs depends upon the position in the network at which β sign is specified; similarly the impulse pattern induced (as an α sign) by a collection of β signs depends upon the set of points specified as recipients. Now the relation between operational languages, which, we assert, depends upon the organism is 'translation' or 'identification of sign sequences' and since any α organism has a localised structural component, position dependence secures this property for any $\beta \rightarrow \alpha$ translation. However, a β organism has a distributed structural component. Hence the property will apply for $\alpha \rightarrow \beta$ translation excepting, perhaps, for the case of a completely distributed β organism.

Identification appears in our system as a process whereby a set of β signs comes to 'mean the same thing to' a certain α organism say G_i^{a} in a region A_1 as a given α sign, say $\mathrm{R}(G_j^{a})$ which is one dynamic component of an α organism G_j^{a} in a region A_2. The phrase, 'meaning the same thing to' G_j^{a} implies that in a λ description the same change of state is induced by a given sequence of β signs as was previously induced or is coincidently induced by the α sign $\mathrm{R}(G_j^{a})$. In other words, a transformation like:

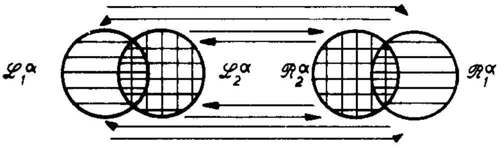


Fig. 5.

is induced either by an α sign or a set of previously incomparable β signs. The same comments apply to the β identification of an α sign.

6. THE EVOLUTION OF ORGANISMS

6.1. *Organisms as decision makers*

In the relatively static picture we have considered so far a group of neurones, say A_1, acts as a decision making system which, at a particular instant, resembles a finite automaton with a transfer function, determined by G_j, and realised as an object in the sense of *3.2* and *3.3* say:

$$F;(X_1^\alpha \otimes Y_1^\alpha) \rightarrow Y_1^\alpha$$

where X_1^α is the set of α signs distinguishable at the incoming fibres of A_1 and Y_1^α is the set of α signs distinguishable at the A_1 outgoing fibres.

But it is important to recall that the substance of the organism is the reproductive process for G_j^α and not the particular neurones that happen to mediate this organisation. The organism sits at A_1 and uses the A_1 neurones as its body, perhaps only for a moment, perhaps indefinitely. The grouping condition of *5.2.* secured a number of positions in the network of neurones where organisms could sit comfortably. Thus a feasible A structure if —>— implies a fibre connection and .... a θ dependence and ○ a neurone, would be

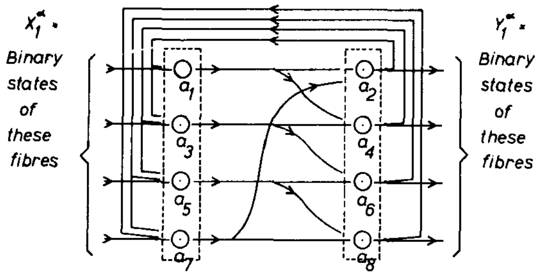


Fig. 6.

and any α organism is an organisation involving loops like $a_1 \leftrightarrows a_2$ or $a_1 \rightarrow a_2 \rightarrow a_3 \rightarrow a_6$ and plastic changes at the synapses of these neurones.

However, the A_1 neurones also mediate parts of β organisms since each neurone is likely to embody several specific receptors, b. We could, though with less practical point, define a distributed β automaton with a transfer function

$$F;(X_B^\beta \otimes Y_B^\beta) \rightarrow Y_B^\beta$$

where X_B^β is the set of β signs appearing in all $b \in B$ and the Y_B^β is the set of β signs produced by the neurones a that embody these specific receptors.

From the engineering point of view any neurone can be regarded as a dual purpose mechanism which includes a number of functional parts (Fig. 7).

The incoming fibres provide impulses which are filtered for their spatial component by an α filter (the θ input to this box is a plausible innovation) and give rise to an α

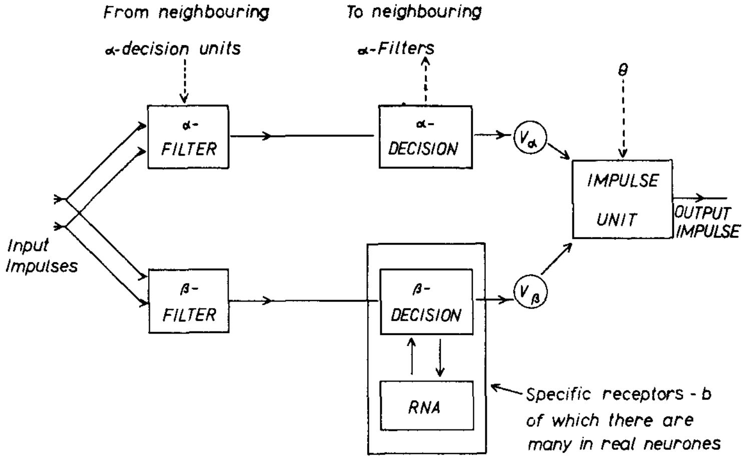


Fig. 7

decision (plausibly a state change that exerts some θ effect upon the neighbouring α filters). The incoming impulses are also filtered by a β filter for their frequency component and give rise to a β decision involving a specific receptor which is shown as a pair of coupled boxes. The neurone produces an output impulse which may be part of either an α sign or a β sign and the decision to emit this impulse depends upon the weight V_α, V_β, given to the α, β, decisions.

6.2. *Interaction with evolving organisms*

Hence, any collection of neurones is potentially the location of an α or a β organism, if it takes part in an α reproducing or a β reproducing process. The common element in these processes is an impulse which is the physical *event* we distinguish as a *sign* in λ.

Suppose we have evidence to suggest that a certain group of neurones *is* the location of an α organism we can interact with it in the operational language α by introducing stimulus and recording electrodes and correlating the spatial or α sign characteristic of its input, with any spatially patterned stimuli we introduce (Fig. 8).

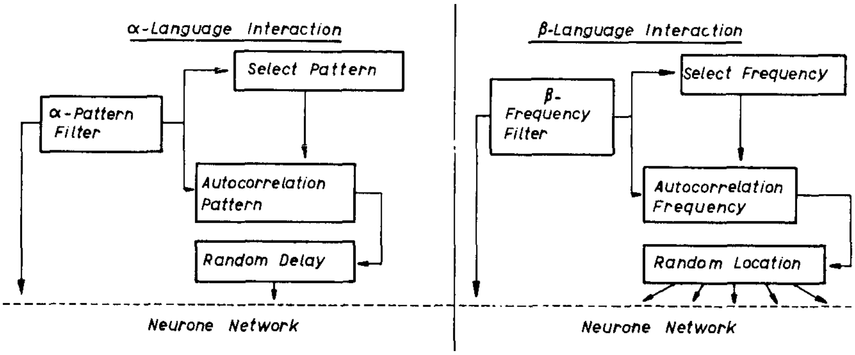


Fig. 8.

At any moment any neurone may be more or less a part of an α or a β organisation. Its status in this respect is chiefly a function of V_α and V_β which, however, we must assume to be variable. Now we have already asserted that a neurone is, in a restricted sense, a θ maximising device and it is no innovation to remark that if the neat boxes did exist (in real neurones, at any rate, they do not), then one of the θ maximising adaptations would be a variation of V_α and V_β.

Hence, suppose we α interact with the neurone group at region A_1 using such a correlation function in our external feedback loop that θ *is* maximised (and if we do not our interaction cannot persist) then the A_1 neurones will be induced to increase V_α and behave as part of an α organism. A similar comment applies in the case of a stable, θ maximising, β interaction when the A_1 neurones will be induced to behave as part of a β organism.

This is the dilemma of any *conversational* method for observing a system. But unless such a definitely α interaction or definitely β interaction is stabilised, the organisms in the system will have α and β aspects. To some extent, we can regain determinacy and the chance to know, for certain, the implication of a sign, by talking, as we have done, about an α organism *identifying* the dynamic component of a β organism as a β sign, or *vice versa*, but this expedient has doubtful value whenever it becomes necessary to admit structural α and β interactions. In fact, the commonest organisms are linguistic hybrids, which is why the system is difficult to observe in any *other* than a conversational way, that is by changing from α to β as necessary, in order to maintain the discourse.

6.3. A self organising system

In *5.2* we defined the internal environment as embodying a set of constraints which would favour the evolution of an organism if it was a self organising system, in other words, only a self organising system would receive an adequate amount of θ.

Consider an α decision maker in the sense of our previous discussion. It will be a self organising system if and only if the rate of change of the redundancy of its behaviour is positive; namely, if and only if $\triangle$ redundancy $(x, y)/\triangle t > 0$.

In a variable internal environment a θ maximising adaptive system can behave as a self organising system in this manner; adaptation will lead to a transfer function which maximises θ, but before this occurs, the environment will change, requiring a different kind of adaptive modification. Thus, in a certain sense, the system is always adapting; and the redundancy of the behavioural outcomes is always, on average, increasing.

This analysis is tempting but spurious because it relies upon redefining the set X and hence redefining the 'system' (Beer, 1961). For a *given* set X (namely in an invariant or stationary environment) the system (with behavioural outcomes $X \otimes Y$) will adapt to maximise θ. The redundancy will increase for a while but will eventually become constant. At this point the system will no longer behave as a self organising system and will no longer be favoured with an adequate value of θ. Since the physical structure of the organism depends upon θ it must *become* a self organising system in order to survive.

6.4. The mechanism of evolution

To become a self organising system the domain of the transfer function, which is a set $X_1 \otimes Y_1$ of behavioural outcomes must be transformed $(X_1 \otimes Y_1) \rightarrow (X_2 \otimes Y_2)$ or $(X_1 \otimes Y_1) \rightarrow T(X_1 \otimes Y_1)$ where T is defined as a transformation that maintains

$$\triangle \text{ redundancy } (x, y,)/\triangle t > 0$$

or equivalently, writing Ξ and $T(\Xi)$ that maintains

$$\triangle \text{ redundancy } (\Xi)/\triangle t > 0$$

Hence, although no system Ξ with an invariant domain can be a *stable* self organising system, the disjunction Ξ^* where

$$\Xi^* = \bigcup_{r=0}^{r=\mathrm{R}} [T^r(\Xi)] \text{ where } r = 1, 2 \, \ldots\ldots\ldots \, \mathrm{R}$$

is a stable self organising system if it exists.

We call Ξ^* the evolution of Ξ and comment that its existence is guaranteed by the physical constraints in networks such as Caianiello's (1961). But the evolution can be either of two kinds.

(1) The set of behavioural outcome *states* is transformed or evolves due to a correlation between the activity of a larger set of neurones. In this case the α organism expands to occupy a larger volume in the network. The mechanism of evolution is transmission of α signs (say from A_1 to a neurone group, harbouring a similar activity like A_2). The result is resonance between the organism located at A_1 and A_2 to yield a system with outcome states $(X_1^\alpha \cup X_2^\alpha) \otimes (Y_1^\alpha \cup Y_2^\alpha)$. Let us call this *regional evolution.*

(2) The set of behavioural outcome *states* is transformed by the inclusion of β states whereby the organism becomes equivalently *capable of identifying β signs* or *an α, β, hybrid*, or in the limit case, a definitely β organism. The mechanism is a change in the weights V_α and V_β (increase in $V_\beta / V_\alpha + V_\beta$) of neurones only some of which will ordinarily be in the group A. The resulting organism will have a set, $(X_1^\alpha \cup \Omega, (X_B^\alpha \otimes (Y_1^\beta \cup \Omega\,(Y_B)$ of behavioural outcomes, but (since X_1 and X_B and Y_1 and Y_B are probably connected with entirely different neurones) this assertion has little practical importance. The *process*, however, *is* important and we shall call it *linguistic evolution.*

6.5. The necessity for evolution

It is not difficult to show that if the evolutionary system is left alone, the organisms will evolve in the regional sense of (1) and by the linguistic evolution of (2). The argument will not be given in this paper but we comment that the result is a consequence of the general principle that one evolving organism will exist in an internal environment increasingly determined by the other organisms of its own species.

6.6. Learning related to evolution

Later in *8.1.* and *8.2.* we shall contend that the regional or the linguistic evolution of an organism has the characteristic required for mechanisms that mediate various kinds of learning behaviour. Before developing the point it will be worth while considering a related case that illustrates the enormous generality of this process.

7. DIGRESSION INTO PHYSICAL ASSEMBLIES

7.1. Solving an undecidable problem

Consider any automaton able to receive evidence from a set of possibilities U and to select actions from a further set of possibilities C. Let the automaton depend upon θ for physical survival, and let it be programmed to maximise the θ available in its environment with a transfer function f; $(U \otimes C) \rightarrow C$ which is often called a decision rule. Suppose that this automaton is presented with an undecidable problem; in other words, given certain evidence $u_0 \in U$ and having selected $c_0 \in C$, the decision rule specifies no change of state, which will increase θ yet u_0 remains invariant and in u_0, c_0, the value of θ is decreasing. Without further specification the automaton is unable to survive. It must remain in u_0, c_0, until it decays.

The organisms we have considered, are momentarily like the automaton. Their strategy for avoiding an undecidable and destructive situation is to *evolve* into another organism which *may* be able to solve the problem.

The present automaton could also be designed with this strategy in which case its response to an undecidable and destructive situation would be $U \rightarrow U^*$, $U \subset U^*$ (to obtain novel evidence on the basis of which to decide) or $C \rightarrow C^*$, $C \subset C^*$ (to select different actions in an attempt to elicit further evidence). The fact that most automata are not designed like this is due to the form of evidence commonly presented, namely the values of one or a limited number of continuous variables. The set of possibilities must still be increased in order to solve the problem but when the environment has this very special structure it is possible to invoke the doctrine of *chance events*, to throw a dice and select amongst the existing set of possibilities *by chance*; for these possibilities define the environment completely.

7.2. Decision in unstable equilibrium

Let us now reduce the choice situation to its most rudimentary form, namely a ball, balanced on a pin, and able to fall in various directions. Suppose the point of the pin is progressively narrowed so that the equilibrium becomes increasingly unstable.

Now this rudimentary situation is 'undecidable' in the sense that at any moment when the ball rests on the pin there is nothing to 'predispose' it to one direction of falling rather than to another; otherwise it would have fallen. As the pin is narrowed, the ball must eventually topple over, when we assume that it is replaced and that the process is repeated.

The event of toppling over can be described in two different ways:

(1) As an event determined by some physical variable or combination of variables such as a breath of wind or a magnetic field. As the pin narrows an increasing number of these variables exert sufficient effect upon the ball to determine its descent. Hence narrowing the pin has the effect of increasing the set of significant variables, that is,

of variables which significantly determine the outcome and of which there are an *indefinitely* large number.

(2) As a matter of chance, for we are ignorant of, or indifferent to the determining variables.

Ordinarily, either (1) or (2) is acceptable. But obviously it would be essential to use (1) and not (2) if the ball could somehow profit from falling in one direction, say X and if being pushed by the breeze determined its descent into X when the pin was less narrow and the system more stable, and if, as a result, the ball adapted to become more sensitive to the breeze than to the other variables. This adaptive process is distinct from adaptations that select a 'value of a variable', such as 'the strength of breeze', which may also occur.

A comparable selection from the states of an indefinitely large number of possible variables takes place in biological evolution to determine, for example, the sense modalities of organisms. Trivially, in the evolutionary system we have considered, or non trivially in the maturation of a biological brain, the selection is made, instead, from forms described in an indefinitely large number of possible languages.

We are now able to relate entities in the evolutionary system to learning mechanisms in a very general way indeed. Having demonstrated the diversity of systems capable of embodying these principles, we need no longer restrict our attention to a pair of operational languages. The evolutionary system, when realised in a physical assembly, has a special and important 'analogue' property (MacKay has pointed out a rather similar property in other systems). Conventionally we are accustomed to analogue devices that select from an indefinite number of values of a variable (the device works with continuous variables). The physically realised evolutionary system selects from an indefinite number of variables or an indefinite number of operational languages.

8. DIFFERENT KINDS OF LEARNING*

8.1. Forms of learning mechanism

The limiting case of a learning mechanism is any system with adaptive properties.

These are a prerequisite of learning or evolution. Consider discourse in α with the evolutionary system we have discussed and suppose that the organism or set of organisms involved in the feedback path of the α signs of the discourse is in a condition where equivalently: (*1*) it must evolve in order to survive; (*2*) it is no longer a stable self organising system; (*3*) it is presented with an undecidable problem; (*4*) it is unstable like the ball on a pin.

Then evolution occurs which will be manifest in the α discourse as a learning behaviour.

If *regional* evolution takes place the possible behaviours are stimulus or response generalisation or the development of a body of related experience; knowledge could be accumulated in terms of different orders of conditional probabilities of events or existing constructs; but no organisation other than the organisations predicated in the a system can occur and hence no essentially novel concept or relation can arise (for even variation in the system which may be interpreted as chance trial learning, induces organisms predicated in the a system). Briefly, then the process of regional evolution implies a mechanism that can account for inductive learning (and, by extending the model to include its response selection) inductive inference in the sense of Braithwaite and Popper (1962). It may also account for change of attention within a field made up of similarly ordered objects. The apparent indeterminacy entailed in the learning process is due to the intrusion of the β signs (or the hybrid character of the evolving organisms) that (in many systems) is also the chief progenitor of variation.

---

* See also the Appendix.

If *linguistic* evolution occurs the resulting behaviour can manifest originality providing that the system is conceivable to whoever indulges in the discourse. a Signs are identified with a set of β signs and *vice versa*. The evolutionary level at which *identification* of a and β signs is deemed to take place depends upon *his* discrimination and is not otherwise demarked from the hybrid interaction which is discounted as irrelevant. So, to use a rather inadequate analogy, some a signs which appear 'hazily valued' in the discourse will at some evolutionary level, assume 'sharp values' as a signs identified with sets of β signs.

The mechanism of linguistic evolution can give rise to novel concepts and relations, for example, the relations involved in constructing the conceptual hierarchy of a hypothetic deductive structure in the sense of Braithwaite and Popper. It can account for attention to previously incomparable objects or entities. Finally, insight is one behavioural consequence of such a system. The flux of a discourse is likely to be, perhaps certain to be, mapped into the interaction of β organisms where it remains as an inaccessible 'hazily valued' β representation until linguistic evolution takes place and sets of β signs are identified with a signs and become 'sharp valued'. Before this occurs, the data may have been modified in its β representation and, if so, we should say that this identification was insightful. Hence, according to our original criteria, learning, in this system, entails linguistic evolution.

8.2. Relation between organisations

In conclusion we should stress the physical characters of this system. Perhaps we have insisted often enough that any organism is designed to survive as a physical entity; that its 'decisions' are made to maintain its fabric and that the abstraction, θ is only a convenient index for whatever energetic conditions are needed to maintain and construct its fabric. On these grounds alone the form of organism that evolves must depend upon the physical assembly that embodies the evolutionary system. Without this material dependence* our conclusions would be limited to systems which literally *had* only an a language and a β language.

But the physical character of the system plays an equally important part in determining the relations, in particular, the similarities, that exist within the system, and in maintaining a correspondence between these and the relations that pertain in an external environment that is built of the same material (MacKay, 1958; Beer, 1961).

---

* The appropriate abstraction for such a materially dependent system is an *infinite* automaton,as discussed, for example, by Loefgren in a recent paper (our evolving organisms are special and rather intractable cases of infinite automata).

In terms of evolving organisms there are relations between individuals and their ancestors and between species (the relations required in *4.4.* and realised by the material constraints). Corresponding to these are relations between forms of sign in a given operational language and between operational languages; physically in the α, β, case, these are realised as relations between spatial patterns *or* between frequencies and relations between spatial patterns *and* frequencies. Finally in discourse there are relations between concepts of the same kind, and concepts of different kinds.

The material dependence of a brain or an artifact suggests that manifestly coherent experience stems from a cohesive physical or intellectual environment of which the brain is a specialised part.

APPENDIX

Comment upon algorithms

In the discussion Professor Rashevsky suggested that the Markovian theory of algorithms could be profitably applied in connection with this learning model. A tentative approach will be outlined, since the method seems to be promising.

We interpret an algorithm as a computing procedure which has conceptual integrity (that, in some sense, Ξ could reveal the algorithms it uses whilst remaining ignorant of their stages of application, though each stage entails some mechanical change). Conceived in this way, an algorithm is a process which reduces Ξ's conceptual uncertainty (given x, y, and x about the next y).

Algorithms η act upon a vocabulary which is some sequential representation, τ of X^a to produce words in $\tau\,(X^a)$ or otherwise to produce words in $\tau\,(Y^a)$. Given (x_0, y_0), Ξ selects some algorithm $\eta \in U(\varphi)$ where all $\eta \in U(\varphi)$, either: (I) satisfy the function f_φ selected, at this instant, from F^a in the sense that $\eta(\tau(x_0)) = f_\varphi(x_0, y_0)$; $x_0 \in X^a$, $y_0 \in Y^a$; or (II) are the first member of a finite sequence of m algorithms

$$\eta_m\,\{\eta_{m-1}\,[.\,.\,.\,.\,.\,.\,.\,\eta\,1\;[\tau(x_0)]\,.\,.\,.\,.\;]\} = f_\varphi(x_0, y_0)$$

where any

$$\eta_{\gamma}\;[\tau(x)] \in \tau(X^a \cup\; Y^a) \text{ for } m \geqslant \gamma \geqslant 1.$$

'Conceptual' uncertainty is envisaged in terms of the average number $v(\varphi)$ of algorithms which are required by Ξ, given x_0, y_0, to satisfy $f_\varphi \in F^a$. In other words, we interpret the average length of sequence of algorithms that is needed to 'solve a problem' as an index of the uncertainty engendered by 'posing' this problem and learning is conceived as a process which maintains the *rate* of *reduction* of v within limits, say, $\xi_1 \geqslant -\Delta v/\Delta t \geqslant \xi_2$; $\xi_1 > \xi_2 > 0$; (hence Ξ is a self organising system). Consider the possibility of interaction between this 'conceptual' level of index and state indices at a mechanical level; in particular, that a *control* is exerted to stabilise $-\Delta v/\Delta t$, by adjustment of these 'mechanical' parameters.

We have distinguished certain underlying *mechanisms* of learning, namely, 'adaptation', 'regional and linguistic evolution'. These occur in a medium or 'environment' (such as a brain or artifact) wherein the rate of activity must, on average, remain constant. Adaptation is a mechanism whereby repeated events ingrain structural patterns; in α spatial distributions of synaptic impedances and in β more or less ramifying, frequency sensitive pathways. Given the mechanical particulars a pair of indices of the degree of adaptation are calculable; say σ_α and σ_β. If σ_α is high then $\varphi \rightarrow \varphi_0$, whilst Ξ is embodied in the particular part of the medium (or internal environment) for which σ_α was calculated.

Regional evolution is a process whereby activity correlations increase the volume of the medium that, at any instant, is embodying Ξ; once again, given the mechanical particulars, an index of correlation is calculable; say ϱ. The higher ϱ the more elaborate the mechanical concomitants of any 'conceptual' event (the application of an algorithm η).

Now *adaptation* implies a restriction upon the range of the available algorithms; its effect is unspecific but, in general, since adaptation implies $\varphi \rightarrow \varphi_0$ and since $v(\varphi_i \cup \varphi_0) > v(\varphi_0)$ the result of adaptation will be a reduction in v and increasing the rate of adaptation will increase $-\Delta v/\Delta t$.

At the conceptual level, decrease in $\mu(\varphi)$ implies a *composition* of algorithms; the selection of $\eta^* [\tau(x)] = \eta_1 [\eta_2 [\tau(x)]]$ in place of η_1 after η_2. Composition increases the number of members of U that have been used at least once. Let $\gamma(\eta)$ be the number of *stages* (which *do* correspond to mechanical events) entailed by the application of η. Commonly, $\gamma(\eta^*) > \gamma(\eta_1) + \gamma(\eta_2)$ as in Rashevsky (1962). We shall interpret the *process* which provides the mechanical elaboration needed to achieve *composition* with *regional* evolution (notice, that, in addition, the regional evolution also shifts the embodiment of Ξ to some other region of the medium with a possibly lower value of σ_α; hence a homeostasis).

Finally, let us identify *linguistic* evolution with the augmentation of a vocabulary. As in Rashevsky (1962) some algorithms, such as those that replicate a word in sequence, cannot act upon any word in a given vocabulary but require additional and distinguished signs which take part in the process (although these signs are unchanged in the final result). Consequently algorithms of this kind act upon an augmented vocabulary, and Dom (U) increases. (Since the augmenting signs must be necessarily *distinct*, mere addition of α signs would not be sufficient.) There is a reasonable and tempting correspondence between *linguistic* evolution involving change in the domain (and/or range) of Ξ and the augmentation of a vocabulary from $\tau(X^\alpha)$ to $\tau(X^\alpha \cup X^\beta)$ and $\tau(Y^\alpha)$ to $\tau(Y^\alpha \cup Y^\beta)$. In this case it is perhaps better to view the appearance of algorithms of this form as a consequence of the reproductive process which we have argued is the mechanical backbone of learning.

Comments on system representation

Let us relate the concepts of *3.4.* to the formal representation of Rosen (1958) and, in particular, deal with the special case considered in detail throughout *5.1.–5.6.* and *6.1.–6.5.*

(1) Suppose a state description of a physical system, Ξ, in the scientific language, λ, yields a set of identified states $z \in Z$. These are partitioned into subsets Z_i (equivalence

classes Z/ϱ induced by the partition ϱ) which correspond to subsystems, $\Xi_i \in \Xi$, $i = 1, 2, \ldots\ldots\ldots n$.

(2) From Rosen (1958) a system is determined by a directed graph in which the directed edges correspond to mappings and the nodes of the graph to the conjunction of the domain, Dom (F), of some mapping, F, with the product set formed from the factors of subsets in the range, Ran (F), of other mappings, all Dom (F) and Ran (F) in Ξ being defined.

Thus $\Xi_i \in F_i$; Dom $(F_i) \rightarrow$ Ran (F_i). If Ξ is the physical embodiment of a communication system, $\Xi_i \in F_i$; Dom $(F_i) \rightarrow Z_i$.

We specify, also, that Z_i is a factor in Dom (F_i), hence:

$$\Xi_i \in F_i;\ [(\overset{*}{Z_j} \otimes \ldots\ldots\ldots \overset{*}{Z}_{j+m}) \otimes Z_i] \rightarrow Z_i$$

where $\overset{*}{Z}_j$, $\overset{*}{Z}_m$, are subsets of Z_j, Z_m, $j = 1,2, \ldots\ldots\ldots n$, not i, and $j + m = j \ldots\ldots n$, not i.

(3) For adaptive subsystems (all subsystems of interest are adaptive) we define a parameter, φ_i, $\varphi_{\max} \geqslant \varphi_i \geqslant 1$, such that φ_i is an index of $f_{\varphi i}$ in a set $F_i = \bigcup_{\varphi i} (f_{\varphi i})$, when $\Xi_i = f_{\varphi i}$; $[(\overset{*}{Z}_j \otimes \ldots\ldots\ldots\ldots \overset{*}{Z}_{j+m}) \otimes Z_i] \rightarrow Z_i$ and Ξ_i is adaptive if, given a stationary sequence, $[x] \subset X_i$ (a redundant condition since all Dom(F) and Ran(F) in Ξ are defined and are connected with similarly specified subsystems) a convergent sequence $[\varphi_i]$ of values of φ_i induces a convergent sequence $[z] \subset Z_i$ so that, after an interval of adaptation, $[z]$, is in a subset $\overset{*}{Z}_i \subset Z_i$.

(4) Adaptation is directed. The convergent sequence of values of φ_i is selected to maximise some variable which we shall identify for the moment with a payoff available to Ξ_i (later we shall give a very different interpretation to this variable).

Formally, let G_i be a function with domain the product set, $Z_i \otimes [z_j \otimes \ldots\ldots z_{j+m}]$ where $n \geqslant m + 1$ which assigns a numerical value to each state in $Z_i \otimes [z_j \otimes \ldots z_{j+m}]$ (each element of this set designates the state of Ξ_i, immediately defined as some element of Z_i, in relation to the states of $m + 1$ other subsystems). Let θ_i denote the average value of G_i over an arbitrary sequence of changes in state.

Adaptation is defined as any process which selects those values of φ_i which maximise θ_i. Since we are not concerned with mechanism so much as a model of mechanism, it is unnecessary to comment upon how this selection is achieved. We take it as a characteristic of the physical medium in which the system is embodied (I) that selection of φ_i (or whatever corresponds to our representation of an adaptive parameter) does occur and (II) that θ_i maximising selections lead to convergent sequences $[\varphi_i]$ of the values of φ_i (hence to an adaptive change).

Finally, we comment that the selective process does entail the idea of 'chance', although only in the weak sense of an 'independent' process which decides between alternative values of φ_i in the absence of sufficient evidence. Consequently, 'chance' is subsumed by the characteristics of the physical medium.

(5) Now Ξ embodies communication systems which will be separated from Ξ in a moment; that is, we shall consider certain information structures needed to maintain the physical stability of Ξ apart from the underlying physically constrained organisation. But a realistic separation rests upon the fact that, in terms of θ_i maximisation,

which is the basis of our stability criteria, communication is worth while, otherwise it cannot, physically, and consistently occur, excepting by accident. We thus assert that G_i is so defined that certain elements of the set $Z_i \otimes (z_j \otimes \ldots\ldots\ldots z_{j+m})$ which can only be consistently, that is, non accidentally, achieved by correlated action on the part of several subsystems, Ξ_j Ξ_m, in conjunction with Ξ_i, have higher θ values for Ξ_i than those which can be achieved independently. G_i induces cooperation which implies interaction and may entail communication.

(6) To derive a communication system from its embodiment in Ξ, we perform an *abstraction* whereby certain subsets of states are identified with *signs*. Now we are concerned with those physical systems, Ξ, which, on abstraction, yield at least a pair of distinct communication systems, A and B, characterised by operational languages α and β. (In the model of *5.1.–5.6.* and *6.1.–6.5.* the abstractions define the spatial pattern and the frequency attributes of physical events.)

Formalise abstraction by a many to one mapping J^α or J^β when $J^\alpha(Z_i) = Y_i^\alpha$, $J^\beta(Z_i) = Y_i^\beta$. Hence a communication subsystem $\Xi_i^\alpha \in A$ is: $\Xi_i^\alpha = f_{\varphi i}^\alpha; X_i^\alpha \otimes Y_i^\alpha \rightarrow Y_i^\alpha$: $f_{\varphi i}^\alpha \in F_i^\alpha = J^\alpha(F_i)$, and where $X_i^\alpha = Y_j^\alpha \otimes \ldots\ldots\ldots\ldots Y_m^\alpha$ (similar definitions apply in the case of $\Xi_k^\beta \in B$). Finally, as a convention, we include one indifferent, or ineffective, value in each Y_i^α or each Y_k^β.

The more familiar *probabilistic* image of a communication subsystem is:

$$\Xi_i^\alpha = P_i^\alpha X_i^\alpha \otimes Y_i^\alpha \rightarrow Y_i^\alpha$$

where P_i^α is a matrix of empirically estimated joint probabilities of utterances, x, y, say $p(x,y)$ and the probabilities $p(f_{\varphi i}) = [p(\varphi_i = 1), p(\varphi_i = 2), \ldots\ldots p(\varphi_i = \varphi_{\max})]$ may be inferred from these empirical estimates.

(7) By previous definition communication in A is possible if and only if the directed graph of A, namely of $J^\alpha(\Xi)$ has cyclic subgraphs; or in B if and only if the directed graph of B, namely $J^\beta(\Xi)$, has cyclic subgraphs. An alternative condition is that Ξ_i^α communicates if and only if:

$$Y_i^\alpha; X_i^{*\alpha} \subset Y_j^\alpha; X_j^{*\alpha} \subset \ldots\ldots\ldots\ldots Y_r^\alpha, \text{ where either (I) } r = i, \text{ or (II) } X_r^{*\alpha} \subset Y_i^\alpha$$

(8) The α vocabulary is $X^\alpha \cup Y^\alpha$; $X^\alpha = \cup_i(X_i^\alpha)$, $Y^\alpha = \cup_i(Y_i^\alpha)$; the β vocabulary is $X^\beta \cup Y^\beta$; $X^\beta = \cup_k(X_k^\beta)$, $Y^\beta = \cup_k(Y_k^\beta)$. The *possible* utterances in α are product pairs $x, y, \in X^\alpha \otimes Y^\alpha$ and in β product pairs, $x, y, \in X^\beta \otimes Y^\beta$ but those which are *achievable* depend upon the functions, $f_{\varphi i}^\alpha$ and $f_{\varphi i}^\beta$ which define the communication subsystems. Any utterance determines a many to one relation between subsystems Ξ_j^α, Ξ_m^α, and some particular subsystem $\Xi_i^\alpha \in A$, the character of which depends upon the form of system graph and the distribution of the values of the φ_i (for the domain of a subsystem is a product set, its range is a factor in a product set). A similar argument applies to $\Xi_k^\beta \in B$.

Most utterances have only a local connotation.

(9) *Legitimate* utterances $x, y, = u \in U^\alpha$ and $x, y, = u \in U^\beta$ (where U^α and U^β are sets of legitimate α and β utterances) 'mean the same thing to' all subsystems $\Xi_i^\alpha \in A$ (or all $\Xi_k^\beta \in B$).

To define 'meaning the same thing' we must refer to the physical embodiment of the communication system; for perhaps the least non trivial definition, that x 'means the same thing to' Ξ^{α}_{i} as it does to Ξ^{α}_{j} is that, in the same conditions, receipt of x induces the 'same *state*' in Ξ^{α}_{i} that it does in Ξ^{α}_{j}.

(10) Let z_0 be a state in Z and let ε be some relation of equivalence in Z. Thus if $z_0 \; \varepsilon \; z$, then z_0, z, are equivalent states and the equivalence class of z_0 with reference to ε is denoted as z_0/ε. Choice of a particular relation ε, though crucial in our argument, is not *determined* in the formal representation; except in the trivial sense that the conceptual framework must not be contradicted by the chosen relation. In particular, if $z_0 \; \varepsilon \; z$, are regarded as the 'same states', signs that 'mean the same thing' must be able to exist.

(11) Consider a pair $x_0 \in X^{\alpha}$, $x_1 \in X^{\alpha}$, with $x_0 \in \mathrm{Dom}(f^{\alpha}_{\varphi i}) \subset X^{\alpha}_{i}$ and similarly with $x_1 \in \mathrm{Dom}(f^{\alpha}_{\varphi i}) \subset X^{\alpha}_{i}$. Let $x_0 \; \pi \; x$, imply that x_0 'means the same thing as' x, if and only if for each $\Xi^{\alpha}_{i} \in A$ and $\Xi^{\alpha}_{j} \in A$ and for the same $z_0 \in Z$ it is true that:

(I) when $(y \in Y^{\alpha}_{i}) = (y \in Y^{\alpha}_{j}) = \hat{y}$, and

(II) when $\varphi_i = \varphi_j = \varphi_0$, φ_0 a value of φ, $f^{\alpha}_{\varphi i}(x_0, \hat{y},) \in z_0/E \cap Z_i$ and that $f^{\alpha}_{\varphi j}(x_1, \hat{y}) \in Z_0/E \cap Z_i$.

(12) We apply the plausible constraint that J^{α} maps $z_0/\varepsilon \cap Z_i$ onto a single member y_a of Y^{α}_{i} (which may be the indifferent element), similarly that Y^{β}_{k} maps $z_0/\varepsilon \cap Z_k$ onto a single member y_b of Y^{β}_{k}.

(13) In this case x_0/π is a *sign* with the common *meaning* of the set z_0/ε. Any $x_a \in x_0/\pi$, $x_b \in x_0/\pi$, may be combined with y_a (if $x_a \in X^{\alpha}$ also) to yield a legitimate utterance $x_a, y_a = u_a \in U^{\alpha}$ and, if $x_b \in X^{\beta}$, to yield a legitimate utterance $x_b, y_b = = u_b \in U^{\beta}$. By 'combined' we intend for x_a 'presented' (given $\hat{y}$ selected by Ξ^{α}_{i} and a suitable value of φ_i) to Ξ^{α}_{i} (or, in the case of the x_b, the same comment with reference to Ξ^{β}_{k} and φ_k, rather than Ξ^{α}_{i} and φ_i).

(14) Sets of signs are *informative* if there exist equivalence classes like z_0/E, z_1/E. . . . and values of φ_i like $\varphi_0, \varphi_1, \ldots\ldots$ Such that if $\varphi_i = \varphi_0$, then $f^{\alpha}_{\varphi i}(x_0, \hat{y},) \; \varepsilon \; z_0/E \cap Z_i$ implying $y_0 \in Y^{\alpha}_{i}$ and if $\varphi_i = \varphi_1$, then instead $f^{\alpha}_{\varphi i}(x_0, \hat{y}) \; \varepsilon \; z_1/E \cap Z_i$ implying $y_1 \in Y^{\alpha}_{i}$; since, in this case, the selection of $y \in Y^{\alpha}_{i}$ indicates the adaptation of Ξ^{α}_{i}.

(15) We are now in a position to specify the composition of the sets of legitimate utterances in the operational languages α and β, namely the sets of product pairs U^{α} and U^{β} which are defined as being disjoint, $U^{\alpha} \cap U^{\beta} = 0$.

Let X/π be the set of equivalence classes x_r/π.

Let Z/ε be the set of equivalence classes z_r/ε.

$$U^{\alpha} = \bigcup_{\substack{x_r/\pi \\ in \\ X/\pi}} \Big\{ \bigcup_{\substack{z_r/E \\ in \\ Z/E}} \{ x_r/\pi \cap [\bigcup_{\Xi^{\alpha}_{i} \in A} (\mathrm{Dom}\,(F^{\alpha}_{i}))] \} \otimes y_r \Big\}$$

Similarly, for U^{β}, if $\Xi^{\alpha}_{i} \in A$ is replaced by $\Xi^{\beta}_{k} \in B$ and if $\mathrm{Dom}(F^{\alpha}_{i})$ is replaced by $\mathrm{Dom}(F^{\beta}_{k})$.

(16) *Coherent* discourse in α or β (incidentally, also *informative* discourse in α or β), consists of sequences of legitimate utterances (this is not, of course, the only discourse in α or β; for some values of φ_i, Ξ^{α}_{i} will, given x_0, produce y such that x_0, y is not in U^{α} and similarly for $\varphi_k, = \Xi^{\beta}_{k}$, and signs not in U^{β}. But insofar as α and β are opera-

tional languages a coherent universe of discourse must exist and communication within it is characterised by sequences $[u_a] \subset U^a$ and $[u_\beta] \subset U^\beta$.

(17) The factors of the x in a coherent sequence of discourse $[u_a] = x_1 \rightarrow \hat{y}_1 \rightarrow \rightarrow x_2 \rightarrow \hat{y}_2 \rightarrow$ must be $y_r \, \varepsilon \, J^a(Z_r/E)$. For since the communication is coherent any selection of $\hat{y}$ from Y^a_i (or, for Ξ^β_k, from Y^β_k) must correspond to some $y_r \, \varepsilon \, J^a(Z_r/E)$ [or in the case of one of the B subsystems to $y_r \, \varepsilon \, J^\beta(Z_r/E)$]. But X^a_i and X^β_k are product sets with factors Y^a_j and Y^β_a; but the only values of these factors which are achieved in coherent discourse are signs or components of signs $\hat{y} = y_r$, as above.

(18) Hence $\hat{y} = J^a(Z_r/E \cap Z_i)$, $\hat{y} \in Y^a_i$, for some r; $\hat{y} = J^\beta(Z_r/E \cap Z_k)$, $\hat{y} \, \varepsilon \, Y^\beta_k$, for some r. Further, if in the directed graph, subsystems Ξ^a_j and Ξ^a_m give an input to subsystem Ξ^a_i; for $x \subset X^a_i$, — $x = J^a(z_a \, . \, . \, . \, . \, . \, . \, . \, z_b)$, where the state $z_a \in Z_a/E \cap Z_j$ and state $z_b \in Z_b/E \cap Z_m$. Thus, using similar arguments for B, we arrive at the structure indicated in Fig. 9 below.

The physical subsystems Ξ_i with behaviours that generate the discourse, are adaptive in the sense of (3) and (4): the values of φ_i converge to a definite value φ_0. Regional evolution entails the stable 'cooperative' coupling of these subsystems (in the sense that the measure of $\mathrm{Dom}\, f_{\varphi_i} \cap \mathrm{Dom}\, f_{\varphi_j}$ increases as a result of adaptation) and it is denoted as '$\Xi_i, \Xi_j \rightarrow \Xi_i \cup \Xi_j$'.

Notice that the utterances of Ξ^a_i are no longer distinct, in λ, from the utterances of Ξ^a_j and become 'joint utterances'. In order to effect separation it would be necessary to discover and attach the name '*i*' or '*j*' as a parameter of each utterance.

Condition (15) and condition (7) imply that in order to preserve the relation $\Xi^a_i = J^a(\Xi_i)$ between Ξ_i and Ξ^a_i as well as $U^a \cap U^\beta = 0$, it is necessary to choose $\mathscr{L}$ and E so that no cyclic subgraph of the system that includes Ξ_i such that $J^a(\Xi_i) = \Xi^a_i \in A$ also includes Ξ_k such that $J^\beta(\Xi_k) = \Xi^\beta_k \in B$. Hence $\Xi_i \cup \Xi_k$ is a prohibited coupling.

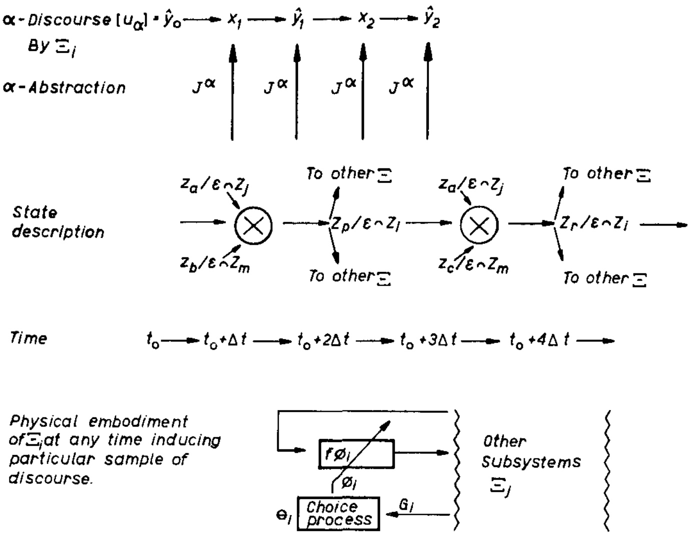

α-Discourse $[u_\alpha]$ • $\hat{y}_0$ → x_1 → $\hat{y}_1$ → x_2 → $\hat{y}_2$
By Ξ_i
α-Abstraction
J^α
J^α
J^α
J^α
State description
$z_a/\varepsilon \cap Z_j$
$z_b/\varepsilon \cap Z_m$
To other Ξ
$z_p/\varepsilon \cap Z_i$
To other Ξ
$z_a/\varepsilon \cap Z_j$
$z_c/\varepsilon \cap Z_m$
To other Ξ
$z_r/\varepsilon \cap Z_i$
To other Ξ
Time
t_0 → $t_0+\Delta t$ → $t_0+2\Delta t$ → $t_0+3\Delta t$ → $t_0+4\Delta t$ →
Physical embodiment of Ξ_i at any time inducing particular sample of discourse.
$f\varnothing_i$
$\varnothing_i$
e_i
Choice process
θ_i
Other Subsystems
Ξ_j

Fig. 9.

(19) There are thus several ways of discussing the system. The most comprehensive refers to the physical embodiment. Here the relations ϱ and ε are preserved and the abstractions J^{α} and J^{β} have a definite form; in the case considered, 'frequency' and spatial pattern attributes of physical events. Next there are the communication systems derived by J^{α} and J^{β} which are related through the equivalence classes induced by ε. Finally, there is the discussion of discourse characterised by a relation π induced by ε; and, speaking in these terms, it may be impossible to preserve the identification between communication systems Ξ_i^{α}, Ξ_k^{β}, and the originally chosen physical subsystems, Ξ_i, Ξ_k.

(20) To preserve the integrity of α and β we need a condition like $U^{\alpha} \cap U^{\beta} = 0$. On the other hand, the possibility of translation depends upon $\alpha\beta$ and $\beta\alpha$ interaction which is formally represented by a many to one mapping Ω_i (with parameter i such that if $X_0^* \subset X^{\beta}$ then *possibly* $\Omega_i(X_0^*) \subset \mathrm{Dom}(F_i^{\alpha})$, and *vice versa* for α to β). But in order to preserve the $\alpha\beta$ distinction if $\Omega_i(X_0^*) \subset \mathrm{Dom}(F_i^{\alpha})$ then it is *not* true that $\Omega_i(X_0^*)\pi x$ for any $x \in X^{\alpha}$. Hence the subsystems we have considered cannot translate coherent discourse.

(21) Translation, in fact, entails the *construction* of a different logical type of statement; hence from the viewpoint of an observer an uncertainty regarding *type* (some care is needed in the interpretation of this remark); for manifestly different types of statement are involved in a discussion of the structure in (19). But translation entails the development of higher *orders* of function. Hence an observer's propositions about the translation are propositions of different and higher type. In particular, his linguistic interpretation of the system leads him to discard his definition of subsystems as ϱ localised entities).

(22) A formal representation of what occurs when a subsystem (at least of the kind we have considered) translates a coherent sequence of discourse leads us to reinterpret our idea of a subsystem or equivalently of the function it computes (the trick involved is that the basic distinction between the subsystems Ξ_i rests upon consideration of the entire structure in Fig. 9 and $\mathscr{L}$ is chosen to satisfy this distinction). So far we have been concerned with utterances U^{α}, and U^{β}, determined by subsystems Ξ_i^{α}, Ξ_j^{α} and Ξ_k^{β}, Ξ_l^{β}, that are characterised by Order (I) functions, F_i^{α}, F_j^{α} and F_k^{β}, F_l^{β}. Now consider the construction of Order (II) (or of higher order) functions determining ,utterances' in $U^{\alpha} \otimes U^{\alpha}$, $U^{\beta} \otimes U^{\beta}$, or since this is not prohibited, $U^{\alpha} \otimes U^{\beta}$.

Denote: (I) x_a in $u_a \in U^{\alpha}$, x_b in $u_b \in U^{\beta}$,

(II) $\xi_a = (x_a, \underline{x}_a)$ in $(u_a, \underline{u}_a) \in U^{\alpha} \otimes U^{\alpha}$

$\xi_b = (x_b, x_b)$ in $(u_b, \underline{u}_b) \in U^{\beta} \otimes U^{\beta}$

$\xi_c = (x_a, \underline{x}_b)$ in $(u_a, \underline{u}_b) \in U^{\alpha} \otimes U^{\beta}$

If π can be interpreted at Order (II) (the assumption that it can is the 'similarity' condition of *3.4.*) then none of these relations are prohibited:

(1) $x_a\, \pi\, \xi_a$, $x_b\, \pi\, \xi_b$, (which are not translations) or

(2) $\xi_a\, \pi\, \xi_c$, $\xi_b\, \pi\, \xi_c$, hence $\xi_a\, \pi\, \xi_b$ when from (1), $x_a\, \pi\, x_b$ which implies the possibility of translation. Now to say '$x_a\, \pi\, x_b$' does contradict the original constraints, imposed upon the system if we interpret this as a proposition of Type I (referring to the Order

(I) functions) but it does not contradict these constraints if it is interpreted as a Type II proposition (in other words if 'mean the same thing as' is now conceived as a relation involving Order (II) functions and the subsystems, whatever these may be, that they characterise). Alternatively, we can define a new relation π^* such that '$x_a \pi^* x_b$' is a Type (I) proposition that is true if and only if the Type (II) proposition '$x_a \pi x_b$' is true.

(23) From (18), regional evolution implies Ξ_i, $\Xi_j \rightarrow \Xi_i \cup \Xi_j$. It can be argued that $\Xi_i \cup \Xi_j$ has a behaviour (and that the corresponding, coupled, communication system generates a discourse) that is determined by a function of higher order than the Order I functions f_{φ_i} and f_{φ_i} that are members of its domain (equivalently the coupling introduces further axioms into the system since the graph is redefined). Linguistic evolution, entailing the 'translation' or 'indentification' of signs, should involve couplings like $\Xi_i \cup \Xi_k$, but these are prohibited by (18). However, the coupling '$[(\Xi_i \cup \Xi_j) \cup (\Xi_k \cup \Xi_l)]$', which is interpreted as a stable, cooperative, association between the coupled systems $\Xi_i \cup \Xi_j$ and $\Xi_k \cup \Xi_l$ is not prohibited. Notice, however, that as in (18) the utterances of Ξ_i^α are inseparable from those of Ξ_j^α or those of Ξ_k^β from Ξ_l^β. Hence, if a mapping, Ω, of utterances is described in the original descriptive framework, in λ, it must carry the names '*i*', '*j*', as a parameter.

(24) Consider, in particular, the special case when a coherent sequence of α discourse μ^α describes a subsystem (since subsystems are identified with functions, it is sufficient that μ^α exhibits the complete behaviour of a subsystem). Then a translation of $\mu^\alpha \rightarrow \mu^\beta$ is a reproduction of the subsystem concerned (notice, (*1*) that this is the case we chiefly examined in *5.1.–5.6.* and *6.1.–6.5.*, and (*2*) the close relation between this case and Rosen's paradox (Rosen, 1959). In order that an α subsystem reproduces it is necessary to invoke a distinct subsystem, here the β image of the α subsystem, and in order to speak about the process to invoke an idea of logical type. In Rosen (1959) the logical type of statements about the system and its *distinct* environment is higher than the original type, of statements about the impossible self *reproducing* system. In the present case the physical correspondent of an α-system is reproduced in the 'environment' provided by the physical correspondent of a β-system and *vice versa*. In terms of α- or β-discourse this entails 'translation'.

(25) Several physical embodiments may be possible; we have considered only a particular evolutionary system, the subsystems of which are characterised by a combination of the Ξ_i, represented, formally, as '$\Xi_i \cup \Xi_j$'.

(26) Now any communicating group of the Ξ_i is trivially such a combination. However, we adopt the reasonable convention of writing '$\Xi_i \cup \Xi_j$' if and only if the correlation between the states of Ξ_i and certain other states of Ξ_j is 1, is such that the interaction between the designated subsystems is not reasonably construed as communication; for no *doubt* is resolved as a result of it.

(27) To achieve $\Xi_i \cup \Xi_j$ the adaptation rules are modified as indicated in *3.4.* and later illustrated in *5.1.–5.6.* and *6.1.–6.5.*

(I) Adaptation is locally irreversible (that is with respect to ϱ and the distinction Z_i and Z_j).

(II) $\theta_i = \theta\ (\Xi_i)$ is an index of survival; we assert that the physical medium that embodies Ξ_i must be maintained at a cost that is measured in terms of $\theta(\Xi_i)$. Hence the $\theta(\Xi_i)$ maximising criterion is a criterion of survival, say $\theta(\Xi_i) > \theta_0$, for survival.

(III) The G_i are so defined that θ is superadditive thus: — $\theta(\Xi_i \, \theta \, \Xi_j) > \theta(\Xi_i) + \theta(\Xi_j)$

(IV) To represent *survival*, the directed graph of Ξ_i is defined if and only if the physical embodiment of Ξ_i survives, namely, if and only if $\theta(\Xi_i) \geqslant \theta_0$.

(28) Finally a replacement rule is needed to avoid the depletion and, ultimately, the disappearance of subsystems. In some pertinent sense the domain and the range of potentially definable sets of functions are created to maintain an average population of n subsystems which differentiate and since θ is superadditive also aggregate by *adaptation.*

SUMMARY

We have considered a cybernetic model in which learning and cognition appear as components of an evolutionary process in a self organizing system. The model illuminates the role of 'distributed' or 'non-localized' functions and it possesses an interesting asymmetry which seems to account for some of the peculiar discontinuities in behaviour associated with 'attention' and 'insight'.

In order to discuss the activity in a self organizing system it is necessary to adopt a descriptive framework applicable in the case when we (acting as observers or experimenters) suffer a structural uncertainty, which we do (though this is often forgotten) in any discussion of intellect. To illustrate the point, we cannot define the representation of a message, like an engineer, with complete certainty. But this does not mean that the model is unconstructable. Indeed, a functioning physical realization of the model exists. Until a few months ago the most likely correlate in the brain for the evolutionary process in this cybernetic model seemed to be a changing mode of activity in resonant neurone circuits. Whilst this interpretation is still plausible, it leads us to look for further physical mechanisms with which the resonant circuits interact. Some recent work from other countries indicates a number of physiological mechanisms acting at a macromolecular level, that may embody the required interactions.

REFERENCES

ANGYAN, A. J., (1958); An analogue model to demonstrate some aspects of neural adaptation. *Mechanisation of Thought Processes.* London, Her Majesty's Stationery Office (p. 933).

ASHBY, W. ROSS, (1956); Design for an intelligence amplifier. *Automata Studies.* C. E. SHANNON AND J. MCCARTHY, Editors. Princeton, N. J., Princeton University Press.

ASHBY, W. ROSS, (1958); The mechanism of habituation. *Mechanisation of Thought Processes.* London, Her Majesty's Stationery Office.

BARBIZET, J., ET ALBARDE, P., (1961); Mémoires humaines et mémoires artificielles. *Concours Méd.,* **6.**

BEER, S., (1961); *Cybernetics and Management.* London, English Universities Press.

BEER, S., (1962); Towards a cybernetic factory. *Principles of Self Organizing Systems.* H. M. Von Foerster and G. W. Zopf, Editors. Proceedings Urbana Symposium. Pergamon Press.

BEURLE, R. L., (1954); Activity in a block of cells capable of regenerating impulses. *R. R. E. Memoranda,* **1042.**

BOK, S. T., (1961); Recognising, learning and interpretation in the vacuole memory theory. *Proceedings 3rd International Congress of Cybernetics, Namur.* Paris, Gauthier-Villars.

BONNER, J. T., (1958); *The Evolution of Development.* Cambridge, University Press.

BRAITHWAITE, R. B., (1962); *Is there a Natural Ordering amongst Theoretical Concepts?* Presidential address to the British Society for the Philosophy of Science, to be published.

BROADBENT, D. E., (1958); *Perception and Communication.* London, Pergamon Press.

BROWN, J., (1958); Information, redundancy and decay of the memory trace. *Mechanisation of Thought Processes.* London, Her Majesty's Stationery Office.

BUSH, R. R., AND MOSTELLER, F., (1955); *Stochastic Models for Learning.* New York, Wiley.

CAIANIELLO, E., (1961); Outline of a theory of thought processes and thinking machines. *Journal of Theoretical Biology*, **2**, 204–235.

CAMERON, S. H., (1960); An estimate of the complexity requisite in a universal decision network. *W.A.D.D. Bionics Symposium.* Washington, A.S.T.I.A.

CARNAP, R., (1947); *Meaning and Necessity. A Study in Semantics and Model Logic.* Chicago.

ESTES, W. K., (1950); Towards a statistical theory of learning. *Psychological Review*, **57**, 94–107.

EYSENCK, H. J., (1956); Reminiscence, drive and personality theory. *Journal of Abnormal and Social Psychology*, **53**, 328–333.

GEORGE, F., (1961); *The Brain as a Computer.* London, Pergamon Press.

GREENE, P. H., (1961); In search of the fundamental units of perception. *Technical Report* **AF 49 (638)–414**. Chicago, United States Air Force.

GUNTHER, G., (1962); Cybernetic ontology. *Proceedings ONR, Armour Research Conference on Self Organising Systems.* Jovitts, Jacobi and Godstein, Editors. Chicago, Spartan Books.

HARMON, L. D., (1959); The artificial neurone. *Science*, **129**, 962.

HARMON, L. D., (1961); Studies with artificial neurones. I. Properties and functions of an artificial neurone. *Kybernetik*, **1**, 89–101.

HEBB, D. O., (1949); *The Organisation of Behaviour.* New York, Wiley.

HULL, C. L., (1952); *A Behaviour System.* New Haven, Conn., Yale University Press.

HYDEN, H., (1960); Paper in: *The Cell, Vol. 4.* J. BRACHET AND A. E. MIRSKY, Editors. New York, Academic Press.

IVAVHEVKO, A., (1961, 1962); *Abtomathka*, and personal communications.

LETTVIN, J. Y., MATTURANA, H. R., AND MCCULLOCH, W. S., (1959); What the frog's eye tells the frog's brain. *Proceedings Institute of Radio Engineers*, **47**, 1940–1951.

MACKAY, D. M., (1951); Mindlike behaviour in artifacts. *British Journal for the Philosophy of Science*, **2**, 105–121.

MACKAY, D. M., (1956); The epistemological problem for automata. *Automata Studies.* C. E. SHANNON AND J. MCCARTHY, Editors. Princeton, N.J., Princeton University Press (pp. 235–251).

MACKAY, D. M., (1958); Operational aspects of intelligence. *Mechanisation of Thought Processes.* London, Her Majesty's Stationery Office.

MASTERMAN, M., (1958); *In various reports*, Cambridge Language Research Unit, Cambridge.

MCCULLOCH, W. S., (1960); The reliability of biological systems. *Self Organising Systems.* M. C. YOVITS AND S. CAMERON, Editors. London, Pergamon Press.

MCCULLOCH, W. S., AND MACKAY, D. M., (1952); Limiting information capacity of a neuronal link. *Bulletin Mathematical Biophysics*, **14**.

MCCULLOCH, W. S., PUTTS, W. R., (1947); How we know Universals. The Perception of Auditory and visual forms, *Bulletin Mathematical Biophysics*, **9**.

MINSKY, M. L., AND SELFRIDGE, O. G., (1960); Learning in random nets. *Information Theory.* C. CHERRY, Editor. London, Butterworth (pp. 335–347).

NAGEL, E., (1949); The meaning of reduction in the natural sciences. *Science and Civilisation.* R. C. STAUFFER, Editor. University of Wisconsin Press.

NAPALKOV, A., (1961); Systematics in the working of the brain and some problems in cybernetics. *Proceedings 3rd International Congress of Cybernetics, Namur.* Paris, Gauthier-Villars.

NAPALKOV, A., BRAINES, S. N., AND SVECHINSKI, V. B., (1959); *Problems in Neurocybernetics.* Moscow, Academy of Sciences.

PAPERT, D., (1960); Redundancy and linear logical nets. *W.A.D.D. Bionics Symposium.* Washington, A.S.T.I.A.

PASK, G., (1958a); Physical analogues to the growth of a concept. *Mechanisation of Thought Processes.* London, Her Majesty's Stationery Office.

PASK, G., (1958b); The growth process inside the cybernetic machine. *Proceedings 2nd International Congress of Cybernetics, Namur.* Paris, Gauthier-Villars.

PASK, G., (1959); A proposed evolutionary model. *Proceedings Urbana Symposium on Principles of Self Organising Systems.* London, Pergamon Press.

PASK, G., (1961a); The cybernetics of evolutionary processes and self organising systems. *Proceedings 3rd International Congress of Cybernetics, Namur.* Paris, Gauthier-Villars.

PASK, G., (1961b); *An Approach to Cybernetics.* London, Hutchinsons.

PASK, G., (1962); Comments on an indeterminacy that characterises a self organising system. *Cybernetics and Neural Processes.* V. BRAITENBERG AND E. CAIANIELLO, Editors. New York, Academic Press.

PASK, G., AND LEWIS, B. N., (1962); *Annual Report on Burroughs Research Corporation Perception Project.* System Research Ltd., to be published.

PAVLOV, I. P., (1927); *Conditional Reflexes.* London, Oxford University Press.

POLONSKY, I., (1961); La cellule vue comme un système cybernétique naturel à la lumière des récents progrès à l'électronique moléculaire. *Proceedings 3rd International Congress of Cybernetics, Namur.* Paris, Gauthier-Villars.

PRINGLE, J. W. S., (1951); On the parallel between learning and evolution. *Behaviour,* **3**, 174.

RASHEVSKY, N., (1960); *Mathematical Biophysics.* New York, Dover.

RASHEVSKY, N., (1962); Mathematical foundations of general biology. *Annals New York Academy of Sciences,* **96**, 4.

ROSEN, R., (1958); Representation of biological systems from the standpoint of the theory of categories. *Bulletin of Mathematical Biophysics,* **20**.

ROSEN, R., (1959); A logical paradox implicit in the notion of a self reproducing automaton. *Bulletin of Mathematical Biophysics,* **21**.

ROSENBLATT, F., (1961); Perceptrons and the theory of brain mechanisms. *Cornell Aeronautical Laboratory Report* No. **VG-1196-G-8**.

SHERRINGTON, C. S., (1948); *The Integrative Action of the Nervous System.* New Haven, Conn., Yale University Press.

SIEGAL, S., (1959); Theoretical models of choice and strategy behaviour; stable state behaviour in the two-choice uncertain outcome situation. *Psychometrika,* **24**, 303–316.

SKINNER, B. F., (1938); *The Behaviour of Organisms; An Experimental Analysis.* New York, Appleton Centory.

UTTLEY, A. M., (1956); Conditional probability machines. *Automata Studies.* C. E. SHANNON AND J. M. MCCARTHY, Editors. Princeton, N. J., Princeton University Press.

VON FOERSTER, H. M., (1960); On self organising systems and their environment. *Self Organising Systems.* M. C. Yovits and S. Cameron, Editors. London, Pergamon Press.

VON FOERSTER, H. M., personal communications.

VON FOERSTER, H. M., AND PASK, G., (1960/61); A predictive model for a self organising system. *Cybernetica.*

WALTER, W. GREY, (1953); *The Living Brain.* London, Duckworth.

WIENER, N., (1961); *Cybernetics.* Cambridge, Mass., Massachusetts Institute of Technology Press; New York, John Wiley.

WIENER, N., (1962); Comments made at the *2nd International Congress of Cybernetic Medicine, Amsterdam, 1962,* and at the *International Spring School of Physics (Cybernetics and Neural Processes), Naples, 1962.*

WILLIS, G. D., (1959); Plastic neurones as memory elements. *Lockheed Report,* **LMSD, 48432.**